PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C12N 15/27, C07K 13/00
A61K 37/02

A1

(11) International Publication Number: WO 92/13075

(43) International Publication Date: 6 August 1992 (06.08.92)

(21) International Application Number: PCT/US92/00433

(22) International Filing Date: 17 January 1992 (17.01.92)

(30) Priority data:
643,502 18 January 1991 (18.01.91) US
757,022 10 September 1991 (10.09.91) US

(71) Applicant: GENETICS INSTITUTE, INC. [US/US]; 87 Cambridge Park Drive, Cambridge, MA 02140 (US).

(72) Inventors: TURNER, Katherine ; 154 Youle Street, Melrose, MA 02176 (US). CLARK, Steven, C. ; 122 Johnson Road, Winchester, MA 01890 (US). JACOBS, Kenneth ; 151 Beaumont Avenue, Newton, MA 02160 (US). HEWICK, Rodney, M. ; 16 Woodcliffe Road, Lexington, MA 02173 (US). GESNER, Thomas, G. ; 27 Shoreline Court, Richmond, CA 94804 (US).

(74) Agent: CSERR, Luann; Genetics Institute, Inc., 87 Cambridge Park Drive, Cambridge, MA 02140 (US).

(81) Designated States: AT (European patent), AU, BE (European patent), CA, CH (European patent), DE (European patent), DK, DK (European patent), ES (European patent), FR (European patent), GB (European patent), GR (European patent), HU, IT (European patent), JP, KR, LU (European patent), MC (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: MEGAKARYOCYTE STIMULATING FACTORS

(57) Abstract

Novel human megakaryocyte stimulating factors (MSFs) capable of stimulating the growth and development of colonies of megakaryocytes, pharmaceutical compositions containing same, and methods for their preparation and use are provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria
AU	Australia
BB	Barbados
BE	Belgium
BF	Burkina Faso
BG	Bulgaria
BJ	Benin
BR	Brazil
CA	Canada
CF	Central African Republic
CG	Congo
CH	Switzerland
CI	Côte d'Ivoire
CM	Cameroon
CS	Czechoslovakia
DE	Germany
DK	Denmark
ES	Spain
FI	Finland
FR	France
GA	Gabon
GB	United Kingdom
GN	Guinea
GR	Greece
HU	Hungary
IE	Ireland
IT	Italy
JP	Japan
KP	Democratic People's Republic of Korea
KR	Republic of Korea
LI	Liechtenstein
LK	Sri Lanka
LU	Luxembourg
MC	Monaco
MG	Madagascar
ML	Mali
MN	Mongolia
MR	Mauritania
MW	Malawi
NL	Netherlands
NO	Norway
PL	Poland
RO	Romania
RU	Russian Federation
SD	Sudan
SE	Sweden
SN	Senegal
SU	Soviet Union
TD	Chad
TG	Togo
US	United States of America

MEGAKARYOCYTE STIMULATING FACTORS

Field

The invention relates generally to a family of novel proteins sharing homologous sequences and biological activities with megakaryocyte colony-stimulating factor (Meg-CSF) and which participate in the differentiation or maturation of megakaryocyte progenitors.

Background

Megakaryocytes are the hematopoietic cells, largely found in the bone marrow but also in peripheral blood and perhaps other tissues as well, that produce platelets (also known as thrombocytes) and subsequently release them into circulation. Megakaryocytes, like all of the hematopoietic cells of the human hematopoietic system, ultimately are derived from a primitive stem cell after passing through a complex pathway comprising many cellular divisions and considerable differentiation and maturation.

The platelets derived from these megakaryocytic cells are critical for maintaining hemostasis and for initiating blood clot formation at sites of injury. Platelets also release growth factors at the site of clot formation that speed the process of wound healing and may serve other functions. However, in patients suffering from depressed levels of platelets (thrombocytopenia) the inability to form clots is the most immediate and serious consequence, a potentially fatal complication of many therapies for cancer. Such cancer patients are generally treated for this problem with platelet transfusions. Other patients frequently requiring platelet transfusions are those undergoing bone marrow transplantation or patients with aplastic anemia.

Platelets for such procedures are currently obtained by plateletphoresis from normal donors. These platelets have a relatively short shelf-life and also expose the patients to considerable risk of exposure to dangerous viruses, such as HIV or hepatitis.

The ability to stimulate endogenous platelet formation in thrombocytopenic patients would reduce their dependence on platelet transfusions and be of great benefit. In addition, the ability to correct or prevent thrombocytopenia in patients undergoing radiation therapy or chemotherapy for cancer would make such treatments safer and possibly permit increases in the intensity of the therapy thereby yielding greater anti-cancer effects.

For these reasons considerable research has been devoted to the identification of factors involved in the regulation of megakaryocyte and platelet production. Such factors are believed to fall into two classes: (1) megakaryocyte colony-stimulating factors (Meg-CSFs), which support the proliferation and differentiation of megakaryocytic progenitors in culture, and (2) thrombopoietic (TPO) factors which support the differentiation and maturation of megakaryocytes in vivo, resulting in the production and release of platelets. [See, e.g., Mazur, E., Exp. Hematol. 15:340-350 (1987).]

Each class of factors is defined by bioassay. Factors with Meg-CSF activity support megakaryocyte colony formation, while factors with TPO activity elicit an elevation in the numbers of circulating platelets when administered to animals. It is not clear how many species of factors exist that have either one or both of these activities. For example, the known factor human IL-3 supports human megakaryocyte colony formation and, at least in monkeys, frequently elicits an elevation in platelet count. However, IL-3 influences hematopoietic cell development in all of the hematopoietic lineages and can be distinguished from the specific regulators of megakaryocytopoiesis and platelet formation, which interact selectively with cells of the

megakaryocytic lineage.

Many different reports in the literature describe factors which interact with cells of the megakaryocytic lineage. Several putative Meg-CSF compositions have been derived from serum [See, e.g., Hoffman, R. et al, J. Clin. Invest. 75:1174-1182 (1985); Straneva, J. E. et al, Exp. Hematol. 15:657-663 (1987); Mazur, E. et al, Exp. Hematol. 13:1164-1172 (1985]. A large number of reports of a TPO factor are in the art. [See, e.g., McDonald, T. P., Exp. Hematol. 16:201-205 (1988); McDonald, T. P., et al, Biochem. Med. Metab. Biol. 37:335-343 (1987); Tayrien, T. et al, J. Biol. Chem. 262: 3262-3268 (1987) and others].

However, biological identification and characterization of Meg-CSF and TPO factors have been hampered by the small quantities of the naturally occurring factors which are present in natural sources, e.g., blood and urine.

The present inventors previously identified a purified Meg-CSF factor from urine described in PCT Publication WO91/02001, published February 21, 1991. This homogeneous Meg-CSF is characterized by a specific activity in the murine fibrin clot assay of greater than $5X10^7$ dilution units per mg and preferably, $2X10^8$ dilution units per mg protein.

There remains a need in the art for additional proteins either isolated from association with other proteins or substances from their natural sources or otherwise produced in homogeneous form, which are capable of stimulating or enhancing the production of platelets in vivo, to replace presently employed platelet transfusions and to stimulate the production of other cells of the lymphohematopoietic system. Such additional proteins are provided by the present invention.

Detailed Description of the Drawings

Figure 1 is a cDNA sequence encoding the MSF precursor containing sequences found in the human urinary Meg-CSF disclosed in PCT Publication WO91/02001, as well as sequences of

other natural and artificial MSFs disclosed herein. Each of the twelve exons has been identified by alternating solid or dashed lines extending from above the first nucleotide in the DNA sequence encoded by that specific exon. The corresponding amino acid sequences appears below each codon.

Figure 2 is a bar graph illustrating the genomic organization of the MSF gene with reference to the number of amino acids encoded by each exon.

Figure 3 is the modified nucleic acid sequence of MSF-K130 which was used to produce the MSF as a fusion protein with thioredoxin in E. coli, as described in Example 5.

Figure 4 illustrates the DNA sequence of the expression plasmid PALTRXA/EK/IL11Δ Pro-581 and the amino acid sequence for the fusion protein used as starting material for production of the thioredoxin/MSF fusion protein described in Example 5.

Detailed Description

The novel family of human megakaryocyte stimulating factors (MSFs) provided by the present invention are protein or proteinaceous compositions substantially free of association with other human proteinaceous materials, contaminants or other substances with which the factors occur in nature. An MSF may be purified from natural sources as a homogeneous protein, or from a selected cell line secreting or expressing it. Mixtures of naturally occurring MSFs may be obtained from natural sources, or from selected cell lines by similar purification techniques. Another class of MSFs are "recombinant or genetically engineered proteins" which are defined herein as naturally occurring and non-naturally occurring proteins prepared by chemical synthesis and/or recombinant genetic engineering techniques, and/or a combination of both techniques. These MSFs may also be provided in optional association with amino acid residues and other substances with which they occur by virtue of expression of the factors in various expression systems. Recombinant or genetically-engineered MSFs of this

invention may further be defined as including a polynucleotide of genomic, cDNA, semisynthetic, or synthetic origin with sequences from the Meg-CSF DNA of Figure 1 which, by virtue of its origin or manipulation: (1) is not associated with all or a portion of a polynucleotide with which it is associated in nature, (2) is linked to a polynucleotide other than that to which it is linked in nature, or (3) does not occur in nature.

The MSFs of the present invention include active fragments and alternatively spliced sequences derived from the DNA and amino acid sequences reported in Figure 1. The nucleotide sequences and corresponding translated amino acids of Figure 1 are continuous in the largest identified cDNA encoding the largest MSF protein, as indicated in the bar graph of Figure 2, which illustrates the genomic organization of the MSF gene with reference to the number of amino acids encoded by each exon. However in Figure 1, each exon has been identified above the DNA sequence encoding that specific exon. While the sequence of Figure 1 is believed to be substantially complete, there may be additional, presently unidentified, exons occurring between Exons VI and IX or following Exon XII, which provide sequence for other members of the MSF family.

The exons of the Meg-CSF gene were identified by analysis of cDNA clones from COS cells transfected with the gene or pieces of the gene or from cDNAs isolated from stimulated human peripheral blood lymphocytes. The first exon, containing the initiating methionine, encodes a classical mammalian protein secretion signal sequence. Exons II through IV contain the amino acid sequences of the original urinary Meg-CSF protein, which most likely terminates in a region between amino acid residues 134 and 205 of Figure 1, based on amino acid sequence data from the native protein. More precisely, the human urinary Meg-CSF protein terminates in the region between amino acid residues 134 and 147. Native, processed Meg-CSF is most likely generated by proteolytic cleavage (endolytic cleavage followed by endolytic and/or exolytic cleavage) at or near this site in

larger precursor molecules containing additional amino acid sequences derived from one or more of Exons V through XII.

During the course of the analysis of the structure of the 18.2 kb "Meg-CSF gene", it was discovered that the primary RNA transcript is spliced in a variety of ways to yield a family of mRNAs each encoding different MSF proteins. In addition, these precursor proteins can be processed in different ways to yield different mature MSF proteins. Thus, a family of MSF's exist in nature, including the Meg-CSF which was isolated from urine from the bone marrow transplant patients. All members of this family are believed to be derived from the 18.2 kb Meg-CSF gene plus a few additional exons, found in the peripheral blood leukocyte cDNA located just downstream from the 3' end of the 18.2 kb gene. The entire 18.2 kb genomic sequence inserted as a NotI fragment in bacteriophage lambda DNA was deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA under accession # ATCC 40856.

This invention also contemplates the construction of "recombinant or genetically-engineered" classes of MSFs, which may be generated using different combinations of the amino acid sequences of the exons of Figure 1. Some of these novel MSFs may be easier to express and may have different biological properties than the native urinary Meg-CSF.

Without being bound by theory, and based on analysis of the naturally occurring Meg-CSF sequence of Figure 1, it is speculated that Exon I is necessary for proper initiation and secretion of the MSF protein in mammalian cells; and that Exon XII is necessary for termination of the translation of the naturally occurring protein. Exons II, III and IV are believed to contain the sequences essential to biological activity of the MSF. Exons V and VI may be related to activity of the factor, but are also implicated in the stability, and folding and processing of the molecule. Exon V and Exon VI are also believed to play a role in the observed synergy of MSF with other cytokines. Alternately spliced forms of MSF cDNAs not

containing Exon V have been observed. Nor has alternative splicing between Exons VI and XII been confirmed. However, preliminary data are consistent with such splicing in the region of Exons VI through XII. Exons V through XII are believed to be implicated in the processing or folding of the appropriate structure of the resulting factor. For example, one or more of Exons V through XII may contain sequences which direct proteolytic cleavage, adhesion, organization of the cellular matrix or extracellular matrix processing. Both naturally occurring MSFs and non-naturally-occurring MSFs may be characterized by various combinations of alternatively spliced exons of Figure 1, with the exons spliced together in differing orders to form different members of the MSF family. At a minimum at least one of the group consisting of Exons II, III and IV and a biologically active fragment thereof is present in a MSF.

Naturally-occurring MSFs may possess at least Exon I, which contains both an initiating methionine necessary for translation and a secretory leader for secretion of the factor from mammalian cells, and one or more additional exons of Figure 1. Of these additional exons, at least one is selected from the group consisting of Exons II, III and IV, and a biologically active fragment thereof. An exemplary MSF of this class includes a protein represented by the spliced-together arrangement of Exons I, II, III. Still another exemplary MSF of this class includes Exons I, III, V and VI.

Other naturally occurring MSFs may possess both Exon I and Exon XII, which latter exon contains a termination codon for translation, and at least one additional exon selected from Exons II, III and IV, and a biologically active fragment thereof. It is speculated that both the initiating Met of Exon I and the termination codon of Exon XII are required to produce an active, properly folded, naturally-occurring MSF in a eukaryotic cell. Thus naturally-occurring MSFs may contain at least Exons I and XII and another exon. An exemplary MSF of

this class includes a protein represented by the spliced-together arrangement of exons selected from Exons I through XII of Figure 1. Still another exemplary MSF of this class includes a protein encoded by the spliced Exons I, II, III, IV, V and XII. Another MSF of this class is formed by spliced together Exons I, II, III, IV and XII. Still another MSF of this class includes the spliced together sequences of Exons I, II, III and XII. Another MSF sequence is formed by spliced together Exons I, III and XII. Yet a further example of an MSF of this class is formed by the spliced together arrangement of Exons I, III, IV and XII.

Another class of naturally occurring MSFs may be characterized by the presence of Exon I, at least one of Exons II, III and IV, or a biologically active fragment thereof, and all of Exons VI through XII. An exemplary MSF of this class includes spliced together Exons I, II, III, IV, and VI through XII. Another MSF of this class is formed by spliced together Exons I, II, III, and VI through XII. Still another MSF sequence is formed from spliced together Exons I, III, and VI through XII. Another MSF sequence of this class includes spliced together Exons I, III, IV and VI through XII.

Still another class of naturally occurring MSFs may be characterized by the presence of Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; and Exons V through XII. An exemplary MSF of this class includes spliced together Exons I, II, III, and V through XII. Another MSF of this class is formed by spliced together Exons I, III, and V through XII. Still another MSF sequence is formed from spliced together Exons I, II, and V through XII. Another MSF sequence of this class includes spliced together Exons I, III, IV and V through XII.

Another class of naturally occurring MSFs may be characterized by the presence of Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; Exon V, and Exons VII through XII. An exemplary MSF of this class

includes spliced together Exons I, II, III, IV, V and VII through XII. Another MSF of this class is formed by Exons I, III, V and VII through XII in a spliced together form. Still another MSF sequence is formed from spliced together Exons I, II, IV, V and VII through XII. Another MSF sequence of this class includes Exons I, III, IV, V and VII through XII spliced together.

Yet another class of naturally occurring MSFs may be characterized by the presence of Exon I, at least one of Exons II, III and IV and a biologically active fragment thereof; at least one of Exons V through XI; and Exon XII. An exemplary MSF of this class includes spliced together Exons I, II, III, IV, V, X and XII. Another MSF of this class is formed by spliced together Exons I, II, III, VIII, IX and XII. Still another MSF sequence is formed from spliced together Exons I, III, VI and XII. Another MSF sequence of this class includes spliced together Exons I, II, IV, V, VII and XII.

For recombinant or genetically engineered MSFs, Exon I may be replaced by a synthetic or heterologous sequence containing an initiating Met and a selected secretory leader designed for use in a selected expression system (hereafter referred to for simplicity as "artificial Exon I"). The natural Exon I may be completely absent for intracellular expression in a bacterial host, such as E. coli. A secretory leader may be selected from among known sequences for secretion of proteins from a variety of host cells. A number of secretory leaders are known for bacterial cells, yeast cells, mammalian cells, insect cells and fungi which may be useful as host cells for expression of a recombinant or genetically-engineered MSF. The construction of an appropriate genetically engineered Exon I sequence containing a secretory leader and initiating Met is within the skill of the art with resort to known sequences and techniques. Thus, one class of recombinant MSFs may be characterized by a genetically-engineered Exon I in place of the naturally occurring Exon I of Figure 1.

Additionally, the termination codon supplied by Exon XII to naturally occurring MSFs may be replaced by inserting into, or after, a selected exon of Figure 1 a termination codon suitable to a selected host cell (hereafter referred to for simplicity as "artificial termination codon"). The construction of an appropriate MSF sequence containing a termination codon is within the skill of the art with resort to known codons for a variety of host cells and conventional techniques. Thus one class of recombinant MSFs may be characterized by the presence of an artificial termination codon.

One class of recombinant MSFs include a naturally-occurring Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; and an artificial termination codon. An example of such an MSF is MSF-K130 and MSF-N141, among others described in detail below.

Another class of recombinant MSFs include an artificial Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; and Exon XII.

Still another class of recombinant MSFs include an artificial Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; and an artificial termination codon.

Another class of recombinant, genetically-engineered MSFs include genetically-engineered Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; and all of Exons V through XII.

Still another class of recombinant MSFs may be characterized by the presence of genetically-engineered Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; and Exons VI through XII.

Another class of recombinant MSFs may be characterized by the presence of genetically-engineered Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; Exon V, and Exons VII through XII.

Yet another class of recombinant MSFs may be characterized

by the presence of genetically-engineered Exon I, at least one of Exons II, III and IV and a biologically active fragment thereof; at least one of Exons V through XI; and an artificial termination codon.

Another class of recombinant MSFs is characterized by genetically-engineered Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; all of Exons V through XI, with an artificial termination codon either inserted into, or added onto a selected last exon of the sequence.

Another class of recombinant MSFs is characterized by genetically-engineered Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; all of Exons VI through XI, with an artificial termination codon.

Another class of recombinant MSFs may be characterized by the presence of native Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof, and all of Exons V through XI, with an artificial termination codon.

Still another class of recombinant MSFs may be characterized by the presence of Exon I, at least one of Exons II, III and IV, and a biologically active fragment thereof; and all of Exons VI through XI, with an artificial termination codon.

Yet another class of recombinant MSFs may be characterized by the presence of Exon I, at least one of Exons II, III and IV and a biologically active fragment thereof; at least one of Exons V through XI; and an artificial termination codon.

A further class of recombinant, genetically-engineered MSFs is characterized by the complete absence of an Exon I. Such MSFs are useful for intracellular expression in bacterial cells, such as E. coli. These MSFs may comprise at least one of Exons II, III and IV and a biologically active fragment thereof; optionally one or more exons from Exons V through XII. In the absence of Exon XII, an artificial termination codon may be inserted into or after the last preferred carboxyl terminal exon. Exemplary MSFs of this invention are MSF-234 and MSF 236

described below in detail.

In another class of naturally-occurring or non-naturally occurring MSFs, either the sequences of Exon VIII and Exon IX will be present together, or neither of these two exons will be present. This is primarily due to frame shifts between these exons and the remaining MSF exons.

Finally an MSF can be made that contains all twelve exons.

While the above described MSF sequence structures will provide for precursor MSFs capable of being processed naturally, or by a host cell expression system, into mature MSF proteins, it is considered that mature, processed forms of the proteins produced in eukaryotic systems will be missing all or part of Exon I. Perhaps the mature proteins may be missing a portion of Exon II as well, in order to remove the leader sequence from the processed form. The processed forms of MSF proteins may also be missing substantial sequences from the carboxyl terminus. For example, sequences from Exons V through XII may be absent in mature, processed MSF proteins. As another example, sequences from Exons VI through XII may be absent in mature, processed MSF proteins. As still another example, sequences from Exons VII through XII may be absent in mature, processed MSF proteins. In such manner human urinary Meg-CSF, an illustrative naturally-occurring MSF, has a mature protein sequence characterized by the presence of Exons II, III and IV in a predominantly homodimeric form.

Selected examples of artificial MSFs were prepared by the following methods. During the analysis of the Meg-CSF gene, a contiguous cDNA was constructed containing Exons I through VI, in which the primary translation product was artificially terminated by inserting artificial termination codons at different positions in Exons IV, V and VI near the point at which the original Meg-CSF was believed to be processed, i.e. the region between amino acid residues 134 and 209. These cDNAs were transfected into COS cells and the resulting supernatants were tested for Meg-CSF activity. Through this process, several

different biologically active MSFs were identified.

One MSF of the present invention is characterized by the DNA sequence extending from nucleotide number 1 of Exon I through nucleotide number 390 of Exon IV, encoding an amino acid sequence which is a continuous fusion in frame extending from amino acid 1 of Exon I through amino acid 130 of Exon IV of the sequence of Figure 1, with a termination codon inserted thereafter. The predicted molecular weight of this MSF is approximately 11.6 kD. On 10-20% gradient sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE), under reducing conditions, a major species of molecular weight of approximately 19 kD has been detected. Under SDS-PAGE non-reducing conditions, the molecular weight ranged from about 20 to about 45 kD. This MSF does not bind heparin under the standard binding conditions of 20 mM tris and pH 7.4. Production and characterization of this molecule, called MSF-K130, is described in detail in Examples 2 and 3.

Upon expression in COS-1 cells, this MSF cDNA sequence produces a mixture of monomeric and homodimeric species. The homodimer has exhibited activity in the fibrin clot assay of Example 10. The MSF expressed by this sequence in mammalian cells approximates the structure and properties of the native human urinary Meg-CSF.

Another MSF of the present invention, called MSF-N141, is characterized by a nucleotide sequence extending from nucleotide number 1 of Exon I through nucleotide number 423 of Exon IV, encoding an amino acid sequence extending from amino acid 1 through amino acid 141 of the sequence of Figure 1 with an artificial termination codon inserted thereafter. The predicted molecular weight of this MSF is approximately 13.2 kD. On 10-20% SDS-PAGE under reducing conditions, a major species of molecular weight of approximately 21 kD has been detected. This MSF binds heparin under standard binding conditions. Upon expression in COS-1 cells, this MSF cDNA sequence produces a mixture of monomeric and homodimeric species. The monomeric

form is the major form secreted by COS-1 cells. The homodimeric form is the major species secreted by CHO cells.

Still another MSF of the present invention, MSF-S172, is characterized by a nucleotide sequence extending from nucleotide numbers 1 of Exon I through 516 of Exon V, encoding an amino acid sequence extending from amino acid 1 through amino acid 172 of the sequence of Figure 1 with an artificial termination codon inserted thereafter. The predicted molecular weight of this MSF is approximately 16.2 kD, and on 10-20% SDS-PAGE under reducing conditions, a major species of molecular weight of approximately 23.5 kD has been detected. This MSF also binds to heparin under standard binding conditions.

A further MSF of the present invention, MSF-R192, is characterized by a nucleotide sequence extending from nucleotide number 1 of Exon I through 576 of Exon V, encoding an amino acid sequence extending from amino acid 1 through amino acid 192 of the sequence of Figure 1 with an artificial termination codon inserted thereafter. The predicted molecular weight of this MSF is approximately 18.4 kD, and on 10-20% SDS-PAGE under reducing conditions, a major species of molecular weight of approximately 27 kD has been detected. This MSF also binds to heparin under standard conditions.

Yet another MSF of the present invention, called MSF-K204, is characterized by a nucleotide sequence extending from nucleotide numbers 1 of Exon I through 612 of Exon VI, encoding an amino acid sequence extending from amino acid 1 through amino acid 204 of the sequence of Figure 1. The predicted molecular weight of this MSF is approximately 19.8 kD. On 10-20% SDS-PAGE under reducing conditions, a major species of molecular weight of approximately 28 kD has been detected. This MSF also binds to heparin under standard conditions.

Still a further MSF of the present invention, called MSF-T208, is characterized by a nucleotide sequence extending from nucleotide numbers 1 of Exon I through 624 of Exon VI, encoding an amino acid sequence extending from amino acid 1 through amino

acid 208 of the sequence of Figure 1 with an artificial termination codon inserted thereafter. The predicted molecular weight of this MSF is approximately 20.4 kD, and on 10-20% SDS-PAGE under reducing conditions, a major species of molecular weight of approximately 29 kD has been detected. This MSF also binds to heparin under standard conditions.

Another MSF of the present invention, MSF-D220, is characterized by a nucleotide sequence extending from nucleotide numbers 1 of Exon I through 660 of Exon VI, encoding an amino acid sequence extending from amino acid 1 through amino acid 220 of the sequence of Figure 1 with an artificial termination codon inserted thereafter. The predicted molecular weight of this MSF is approximately 21.6 kD, and on 10-20% SDS-PAGE under reducing conditions, a major species of molecular weight of approximately 30 kD has been detected. This MSF also binds to heparin under standard conditions.

Additional MSFs of this invention include MSF-T133 (including nucleotides 1 through 399 of Figure 1 and encoding amino acids 1 through 133 with an artificial termination codon inserted thereafter), MSF-R135 (Figure 1 nucleotides 1 through 405 encoding amino acids 1 through 135 with an artificial termination codon inserted thereafter), MSF-P139 (Figure 1 nucleotides 1 through 417 encoding amino acids 1 through 139 with an artificial termination codon inserted thereafter), MSF-K144 (Figure 1 nucleotides 1 through 432 encoding amino acids 1 through 144 with an artificial termination codon inserted thereafter), MSF-K147 (Figure 1 nucleotides 1 through 441 encoding amino acids 1 through 147 with an artificial termination codon inserted thereafter) and MSF-E157 (Figure 1 nucleotides 1 through 471 encoding amino acids 1 through 157 with an artificial termination codon inserted thereafter).

Although in all of the above-described MSFs, the amino and carboxy termini of each MSF is defined precisely, it is to be understood that addition or deletion of one or several amino acids (and consequent DNA coding region) from either end of any

of the MSFs (or from either end of any of the exons forming the spliced MSFs) is not likely to significantly alter the properties of the particular MSF. Such truncated MSFs which also retain MSF biological activities are also encompassed by this disclosure. The deliberate insertion of artificial termination codons at other positions in the MSF sequences can provide other members of the MSF family.

The alternatively spliced MSFs of the present invention are characterized by amino acid sequences containing at least two exons and less than twelve exons of Figure 1 as described above, which exons are spliced together in various arrangements. Several representative "alternatively-spliced" naturally occurring MSF sequences have been identified by polymerase chain reaction (PCR) of cDNA prepared from various cell lines. The sequences of these MSFs were confirmed by hybridization to oligonucleotides spanning exon junctions, molecular weight of PCR fragments, and by DNA sequence in one case. A second method of obtaining MSF sequences involved natural isolation of cDNAs from a HeLa cDNA library. The molecular weights of these MSFs were calculated.

In the PCR technique, the primers were contained within Exons I and VI and were designed to PCR between these exons. Therefore, these exons may all be present in these MSFs. Alternatively, no exon from exon VI through XII may be present. Still alternatively one or more of Exon VI through XII may be present in these representative alternately spliced MSFs.

For example, the 5' end of one such MSF, called MSF-136 (containing Exons I, III and VI), identified by PCR, is characterized by a contiguous amino acid sequence containing amino acid 1 to 25 of Exon I (nucleotides 1 through 76 of Figure 1) joined in frame to amino acid 67 to 106 of Exon III (nucleotides 200 through 319, joined in frame to amino acid 200 to about 250 of Exon VI (nucleotides 598 through about 748. Although not identified by a PCR primer, additional 3' sequence may be present in this MSF, as in each of the below described

PCR-identified sequences. This 5' MSF sequence has been detected in the cDNA of the following cell lines: the osteosarcoma cell line U20S (ATCC No. HTB96), the small cell lung carcinoma cell line H128 (ATCC No. HTB120), the neuroblastoma cell line SK-N-SH (ATCC No. HTB11), the neuroblastoma cell line SK-N-MC (ATCC No. HTB10), the erythroleukemia cell lines OCIM1 and OCIM2, the erythroleukemia cell line K562 (ATCC No. CCL243) following culture in the presence or absence of phorbol myristate acetate, the hepatoma cell line HEPG2 (ATCC No. HB8065) and in stimulated peripheral blood leukocytes from normal volunteers (PBLs). Its presence indicates that a naturally-occurring alternately spliced MSF may comprise Exons I, III, VI and optionally one or more of Exons VII through XII. An artificial MSF mimicking this structure may have an artificial termination codon inserted within or after Exon VI.

Another PCR-identified 5' MSF sequence, containing Exons I, II, III and VI and called MSF-1236, is characterized by a contiguous amino acid sequence containing amino acid 1 to 25 (nucleotides 1 through 76) of Exon I joined in frame to amino acid 26 to 66 (nucleotides 77 through 199) of Exon II, joined in frame to amino acid 67 to 106 (nucleotides 200 through 319) of Exon III, joined in frame to amino acid 200 to about 250 (nucleotides 598 through about 748) of Exon VI. This 5' MSF sequence has been detected by PCR analysis of the following cell lines: U20S, H128, SK-N-SH, SK-N-MC, the neuroglioma epithelial-like cell line H4 (ATCC No. HTB148), OCIM1, OCIM2, K562, K562 in the presence of PMA, the erythroleukemia cell line HEL (ATCC No. TIB180) in the presence of PMA, OCIM2, HEPG2 and stimulated PBLs. The presence of this MSF-1236 indicates that a naturally-occurring alternately spliced MSF may comprise Exons I, II, III, VI and optionally one or more of Exons VII through XII. A recombinant MSF mimicking this structure may have an artificial termination codon inserted within or after Exon VI.

Still another MSF according to this invention is

characterized by a contiguous amino acid sequence containing amino acid 1 to 25 (nucleotides 1 through 76) of Exon I joined in frame to amino acid 26 to 66 (nucleotides 77 through 199) of Exon II, joined in frame to amino acid 67 to 106 (nucleotides 200 through 319) of Exon III, joined in frame to amino acid 107 to 156 (nucleotides 320 through 469) of Exon IV, joined in frame to amino acid 200 to 1140 (nucleotides 598 through 3421) of Exon VI. This MSF-12346 has been detected by PCR analysis of the following cell lines: U20S, SK-N-SH, SK-N-MC, OCIM1 in the presence of PMA, K562 in the presence of PMA, HEPG2 and stimulated PBLs. The presence of this indicates that a naturally-occurring alternately spliced MSF may comprise Exons I, II, III, IV, VI and optionally one or more of Exons VII through XII. A recombinant MSF mimicking this structure may have an artificial termination codon inserted within or after Exon VI.

Another MSF sequence of this invention may include MSF-1234, characterized by a contiguous amino acid sequence containing amino acid 1 to 25 (nucleotides 1 through 76) of Exon I (signal peptide) joined in frame to amino acid 26 to 66 (nucleotides 77 through 199 of Exon II, joined in frame to amino acid 67 to 106 (nucleotides 200 through 319) of Exon III, joined in frame to amino acid 107 to 156 (nucleotides 320 through 469) of Exon IV. This sequence optionally has a 3' sequence comprising one or more of Exons V through XII. This sequence may also contain an artificial termination codon inserted within or after any selected C-terminal exon.

Still another MSF sequence, MSF-134, is characterized by a contiguous amino acid sequence containing amino acid 1 to 25 (nucleotides 1 through 76) of Exon I (signal peptide) joined in frame to amino acid 67 to 106 (nucleotides 200 through 319) of Exon III, joined in frame to amino acid 107 to 156 (nucleotides 320 through 469) of Exon IV. This sequence optionally has a 3' sequence comprising one or more of Exons V through XII. This sequence may also contain an artificial termination codon

inserted within or after any selected C-terminal exon.

Two examples of MSFs that may be useful for bacterial intracellular expression include MSF-234, characterized by a contiguous amino acid sequence containing amino acid 26 to 66 (nucleotides 77 through 199) of Exon II joined in frame to amino acid 67 to 106 (nucleotides 200 through 319) of Exon III, joined in frame to amino acid 107 to 156 (nucleotides 320 through 469) of Exon IV; and MSF-236,characterized by a contiguous amino acid sequence containing amino acid 26 to 66 (nucleotides 77 through 199) of Exon II joined in frame to amino acid 67 to 106 (nucleotides 200 through 319) of Exon III, joined in frame to amino acid 200 to 1140 (nucleotides 598 through 3421) of Exon VI. These sequences each optionally may have a 3' sequence comprising one or more of Exons V through XII. These sequences may also contain an artificial termination codon inserted within or after any selected C terminal exon.

It is further contemplated by the present invention that other MSFs which may be characterized by having MSF biological activities and which may be useful as research agents, diagnostic agents or as therapeutic agents, include factors having other combinations and arrangements of two or more of the exons identified in Figure 1. The splicing of the exons to form recombinant MSFs may be accomplished by conventional genetic engineering techniques or chemical synthesis, as described herein.

Additionally, analogs of MSFs are included within the scope of this invention. An MSF analog may be a mutant or modified protein or polypeptide that retains MSF activity and preferably has a homology of at least about 50%, more preferably about 70%, and most preferably between about 90 to about 95% to human urinary Meg-CSF. Still other MSF analogs are mutants that retain MSF activity and preferably have a homology of at least about 50%, more preferably about 80%, and most preferably between 90 to 95% to MSF-K130 and the other truncated MSFs. Typically, such analogs differ by only 1, 2, 3, or 4 codon

changes. Examples include MSFs with minor amino acid variations from the amino acid sequences of native or recombinant Meg-CSF, or any of the above-described MSFs, in particular, conservative amino acid replacements. Conservative replacements are those that take place within a family of amino acids that are related in their side chains. Genetically encoded amino acids are generally divided into four families: (1) acidic = aspartate, glutamate; (2) basic = lysine, arginine, histidine; (3) non-polar = alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan; and (4) uncharged polar = glycine, asparagine, glutamine, cystine, serine, threonine, tyrosine. Phenylalanine, tryptophan, and tyrosine are sometimes classified jointly as aromatic amino acids. For example, it is reasonable to expect that an isolated replacement of a leucine with an isoleucine or valine, an aspartate with a glutamate, a threonine with a serine, or a similar conservative replacement of an amino acid with a structurally related amino acid will not have a major effect on the MSF activity, especially if the replacement does not involve an amino acid at the active site of the MSF.

The MSFs of this invention may form monomers or homo- or hetero-dimers when expressed in suitable expression systems, due to the presence of cysteine-rich sequences in the exons. As indicated above, two specific homodimeric forms have been identified, namely the MSF-K130 characterized by the sequence of amino acid 1 through 130 of Figure 1, and the MSF-N141 characterized by the sequence of amino acid 1 through 141 of Figure 1. These homodimeric forms were found as abundant forms of these proteins. However, these proteins existed in mixtures of other dimeric and monomeric forms.

Other MSFs of this invention are predominantly monomers rather than mixtures, such as the MSFs characterized by the sequence of amino acids 1 through 209 of Figure 1, or amino acids 1 through 172 of Figure 1, among others.

MSFs of the present invention may act directly or

indirectly on megakaryocyte progenitor cells and/or megakaryocytes. MSFs may act directly on accessory cells, such as macrophages and T cells, to produce cytokines which stimulate megakaryocyte colony formation. Specifically, MSFs display megakaryocyte colony stimulating activity. Another MSF activity is the promotion of megakaryocyte maturation. The active MSF compositions of the present invention have biological activity in the murine fibrin clot megakaryocyte colony formation assay. For example, the MSF characterized by the amino acid sequence amino acid 1 of Exon I through amino acid 130 of Exon IV (MSF-K130) has a specific activity of greater than approximately 1 X 10^7 dilution units/mg protein.

MSFs may also be used in synergy with other cytokines. For example, MSFs also display enhancement of IL-3-dependent megakaryocyte colony formation. MSFs also display enhancement of steel factor-dependent megakaryocyte colony formation. Together, these cytokines, IL-3 and steel factor, have been shown to stimulate increased megakaryocyte colony formation in vitro. In addition, IL-3 has been shown to elevate the level of platelets in non-human primates in vivo.

It is contemplated that all MSFs encoded by the combinations of sequences selected from Exons I through XII as reported above will have MSF biological activity, for example, activity in the murine fibrin clot assay, either alone or in combination with other cytokines. All modified or mutant MSF peptides or polypeptides of this invention, including the spliced forms of MSF, may be readily tested for activity in the megakaryocyte fibrin clot assay, either alone or in combination with other known cytokines including IL-3, steel factor or GM-CSF. Other cytokines which may be useful in combination with the MSFs of this invention include G-CSF, M-CSF, GM-CSF, IL-1, IL-4, erythropoietin, IL-6, TPO, IL-11, LIF, urinary Meg-CSF, IL-7 and IL-9.

These MSFs may also have biological or physiological activities in addition to the ability to stimulate the growth

and development of megakaryocyte colonies in culture in the assay using murine bone marrow target cells. In the murine fibrin clot megakaryocyte colony formation assay, an MSF composition of the present invention stimulates the growth of multiple colony types, but at least 50% of the colonies are pure megakaryocytic or mixed lineage colonies having significant numbers of megakaryocytes. The exact composition of colony types may vary with different assay conditions (fetal calf serum lots, etc). Among the megakaryocyte-containing colonies, typically 50-70% are pure megakaryocytic in composition. In some cases, the particular MSF may not by itself stimulate megakaryocyte colony formation, but rather may enhance megakaryocyte colony formation supported by other factors, such as IL-3 or steel factor; or it may synergize with other factors, such as IL-11, which alone is not capable of supporting megakaryocyte colony formation in the fibrin clot assay.

In the murine agar megakaryocyte colony formation assay, an MSF of the present invention will stimulate colonies of megakaryocytes. Similarly, in the human plasma clot megakaryocyte colony formation assay, an MSF of the present invention will stimulate colonies of megakaryocytes.

It is presently anticipated that maximal biological activities of these MSFs <u>in vitro</u> may be achieved by activating the factors with acid, or denaturing conditions in SDS-PAGE, or by reverse phase high pressure chromatography (RP-HPLC). With both the native urinary protein and the recombinant MSF-K130, an increase in the number of units of activity has been routinely detected after SDS-PAGE and RP-HPLC.

The present invention also encompasses MSF-encoding DNA sequences, free of association with sequences and substances with which the DNA occurs in natural sources. These DNA sequences, including the sequences reported in Figure 1 and identified above, code for the expression for MSF polypeptides. These sequences, when expressed in mammalian cells, yield precursor MSFs which are processed in the mammalian cells to

yield functional proteins. Similar processing is expected to be seen in other non-mammalian expression systems.

Examples of MSF DNA sequences may include a DNA sequence comprising nucleotides 1 through 390 of Figure 1. Another MSF DNA sequence comprises nucleotides 1 through 423 of Figure 1. Another MSF DNA sequence comprises nucleotides 1 through 516 of Figure 1. Yet another example of an MSF DNA sequence comprises nucleotides 1 through 576 of Figure 1. Still a further illustration of an MSF DNA sequence comprises nucleotides 1 through 612 of Figure 1. An additional MSF DNA sequence comprises nucleotides 1 through 624 of Figure 1. An MSF DNA sequence may also comprise nucleotides 1 through 660 of Figure 1. Additional MSF DNA sequences may also comprise nucleotides 1 through 399, nucleotides 1 through 405, nucleotides 1 through 417, nucleotides 1 through 432, nucleotides 1 through 441 and nucleotides 1 through 471 of Figure 1.

Other MSF DNA sequences include the 5' sequences of certain alternately spliced MSFs, such as a sequence comprising nucleotides 1-76 of Figure 1 fused in frame to nucleotides 200-319 of Figure 1, fused in frame to nucleotides 598-748 of Figure 1. Another such 5' DNA sequence comprises nucleotides 1-319 of Figure 1 fused in frame to nucleotides 598-748 of Figure 1. Still another DNA sequence comprising nucleotides 1-469 of Figure 1 fused in frame to nucleotides 598-748 of Figure 1. Another MSF DNA sequence comprises nucleotides 1-76 of Figure 1 fused in frame to nucleotides 200-319 of Figure 1, fused in frame to nucleotides 598-748 of Figure 1. Still another DNA sequence ends from nucleotides 1 through 469 of Figure 1. Another MSF DNA sequence comprises nucleotides 1 to 76 of Figure 1 fused in frame to nucleotides 200 through 469 of Figure 1.

Other MSF DNA sequences which encode homo- or hetero-dimers of the above-described MSF DNA sequences or DNA sequences encoding a biologically active fragment of such sequences are also included in this invention. Similarly an allelic variation of the MSF DNA sequences, and a DNA sequence capable of

hybridizing to any of MSF DNA sequences, which encodes a peptide or polypeptide having activity in the fibrin clot assay are also encompassed by this invention.

It is understood that the DNA sequences of this invention which encode biologically active human MSFs may also comprise DNA sequences capable of hybridizing under appropriate conditions, or which would be capable of hybridizing under said conditions, but for the degeneracy of the genetic code, to an isolated DNA sequence of Figure 1 or to active MSFs formed by alternate splicing of two or more exons of Figure 1 as described above. These DNA sequences include those sequences encoding all or a fragment of the above-identified exon peptide sequences and those sequences which hybridize under stringent or relaxed hybridization conditions [see, T. Maniatis et al, Molecular Cloning (A Laboratory Manual), Cold Spring Harbor Laboratory (1982), pages 387 to 389] to the MSF DNA sequences. Stringent hybridization is defined as hybridization in 4XSSC at 65°C, followed by a washing in 0.1XSSC at 65°C for an hour. Alternatively, stringent hybridization is defined as hybridization in 50% formamide, 4XSSC at 50°C.

DNA sequences which hybridize to the sequences for an MSF under relaxed or "non-stringent" hybridization conditions and which code for the expression of MSF peptides having MSF biological properties also encode novel MSF polypeptides. Non-stringent hybridization is defined as hybridization in 4XSSC at 50°C or hybridization with 30-40% formamide at 42°C. For example, a DNA sequence which shares regions of significant homology, e.g., Exons II, III or IV, and/or glycosylation sites or disulfide linkages, with the sequences of MSF and encodes a protein having one or more MSF biological property clearly encodes an MSF polypeptide even if such a DNA sequence would not stringently hybridize to the MSF sequences. The DNA sequences of this invention may include or contain modifications in the non-coding sequences, signal sequences or coding sequences based on allelic variation among species, degeneracies of the genetic

code or deliberate modification. Allelic variations are naturally-occurring base changes in the species population which may or may not result in an amino acid chang . Degeneracies in the genetic code can result in DNA sequences which code for MSF polypeptides but which differ in codon sequence. Deliberate modifications can include variations in the DNA sequence of MSF which are caused by point mutations or by induced modifications to enhance the activity, half-life or production of the polypeptides encoded thereby. All such sequences are encompassed in the invention. Utilizing the sequence data in Figure 1 and the exon combinations described above, as well as the denoted characteristics of MSF, it is within the skill of the art to modify DNA sequences encoding an MSF and the resulting amino acid sequences of MSF by resort to known techniques.

Modifications of interest in the MSF sequences may include the replacement, insertion or deletion of a selected nucleotide or amino acid residue in the coding sequences. For example, the structural gene may be manipulated by varying individual nucleotides, while retaining the correct amino acid(s), or the nucleotides may be varied, so as to change the amino acids, without loss of biological activity. Mutagenic techniques for such replacement, insertion or deletion, e.g., in vitro mutagenesis and primer repair, are well known to one skilled in the art [See, e.g., United States Patent No. 4,518,584]. The encoding DNA of a naturally occurring MSF may be truncated at its 3'-terminus while retaining its biological activity. A recombinant, genetically-engineered MSF DNA sequence may be altered or truncated at both its 3' and 5' termini while retaining biological activity. It also may be desirable to remove the region encoding the signal sequence, and/or to replace it with a heterologous sequence. It may also be desirable to ligate a portion of the MSF sequence to a heterologous coding sequence, and thus to create a fusion peptide with the biological activity of MSF.

Specific mutations of the sequences of an MSF polypeptide may involve modifications of a glycosylation site. The absence of glycosylation or only partial glycosylation results from amino acid substitution or deletion at any asparagine-linked glycosylation recognition site or at any site of the molecule that is modified by addition of O-linked carbohydrate. An asparagine-linked glycosylation recognition site comprises a tripeptide sequence which is specifically recognized by appropriate cellular glycosylation enzymes. These tripeptide sequences are either Asn-X-Thr or Asn-X-Ser, where X can be any amino acid except proline. For example, such a site can be found in the cDNA illustrated in Figure 1 at amino acids #206-#208. A variety of amino acid substitutions or deletions at one or both of the first or third amino acid positions of a glycosylation recognition site (and/or amino acid deletion at the second position) results in non-glycosylation at the modified tripeptide sequence. Expression of such altered nucleotide sequences produces variants which are not glycosylated at that site.

Other analogs and derivatives of the sequence of an MSF which would be expected to retain MSF activity in whole or in part may also be easily made by one of skill in the art given the disclosures herein. One such modification may be the attachment of polyethylene glycol (PEG) onto existing lysine residues in an MSF sequence, as taught in United States Patent No. 4,904,584, which is incorporated herein by reference. Alternatively, the insertion of one or more lysine residues or other amino acid residues that can react with PEG or PEG derivatives into the sequence by conventional techniques may enable the attachment of PEG moieties. Existing cysteines may be used according to techniques taught in PCT Publication WO90/12874.

In addition to the above, other open reading frames (ORFs) or structural genes encoding MSFs may be obtained and/or created from cDNA libraries from other animal cell sources. For

example, a murine MSF genomic clone and several partial MSF cDNA clones have been isolated by the inventors.

A naturally occurring MSF of this invention may be obtained as a single homogeneous protein or mixture of various alternately spliced MSF proteins and purified from natural sources. Among such natural sources are human urine or stimulated PBLs, other mammalian cell sources producing the factors naturally or upon induction with other factors from cell lines. The DNA of such MSFs may also be obtained and purified from natural sources.

To isolate and purify the naturally-occurring MSFs from natural sources, the purification technique comprises the following steps which are described in more detail in Example 1 below. The example and the following summary illustrate the purification for an exemplary naturally-occurring MSF, human urinary Meg-CSF, which is isolated from human urine. For the urinary Meg-CSF, the purification includes concentrating pooled bone marrow transplant patient urine through an Amicon YM-10 filter. The concentrated urine is passed through an anion exchange chromatographic column and the flow-through is bound onto a cation exchange chromatographic column. The urinary protein eluate is then subjected to pooling, dialyzing and heating and is applied to a lectin affinity chromatographic column. This eluate is then dialyzed and applied to a cation exchange fine performance liquid chromatography (FPLC) column. Finally this eluate is applied through three cycles of reverse phase high pressure liquid chromatography (HPLC) using different solvent systems for each HPLC run.

According to this purification scheme, batches with the highest levels of MSF in the murine fibrin clot assay, described below, are selected for further purification at the semi-preparative scale (between 30 and 100 liters urine equivalent) to maximize recovery and yield. Thus a homogeneous MSF, native Meg-CSF, may be obtained by applying the purification procedures described in Example 1 to human urine or other sources of human

MSF, e.g., activated PBLs.

Other tissue sources and cell lines from which naturally occurring MSFs may be isolated include HeLa cell lines, e.g. ATCC #098-AH2, and bone marrow cell lines, such as murine bone marrow cell line, FCM-1 [Genetics Institute, Inc., Cambridge, MA], osteosarcoma cell line U20S, small cell lung carcinoma H128, neuroblastoma SK-N-SH, neuroblastoma SK-N-MC, neuroglioma epithelial-like cell line H4, erythroleukemia cell line OCIM1 and OCIM2, erythroleukemia cell line K562 in the presence of PMA, erythroleukemia cell line HEL in the presence of PMA, and hepatoma cell line HEPG2. Procedures for culturing a cell source which may be found to produce an MSF are known to those of skill in the art. The MSF proteins and the DNA sequences encoding MSFs of this invention can be produced via recombinant genetic engineering techniques and purified from a mammalian cell line which has been designed to secrete or express the MSF to enable large quantity production of pure, active MSFs useful for therapeutic applications. The proteins may also be expressed in bacterial cells, e.g., E. coli, and purified therefrom. The proteins may also be expressed and purified in yeast cells or in baculovirus or insect cells. Alternatively, an MSF or active fragments thereof may be chemically synthesized. An MSF may also be synthesized by a combination of the above-listed techniques. Suitable techniques for these different expression systems are known to those of skill in the art.

To produce a recombinant MSF, the DNA sequence encoding the factor can be introduced into any one of a variety of expression vectors to make an expression system capable of producing an MSF or one or more fragments thereof in a selected host cell.

The DNA sequences of the individual exons may be obtained by chemical synthesis or may be obtained from the following two deposits by standard restriction enzyme subcloning techniques or by the polymerase chain reaction (PCR) using synthetic primers for each exon based on the nucleotide sequences of Figure 1.

Two genomic clones containing Meg-CSF sequences which may be used as sources of the MSF sequences have been deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland 20852, USA in accordance with the requirements of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure on August 3, 1990.

An approximately 12 kb genomic fragment (referred to as Meg Kpn-SnaBI) containing the sequences spanning Exon I through part of Exon VI [See Table 2, the 5' KpnI site to the 3' SnaBI site] in an E. coli plasmid was given the accession number ATCC 40857. As described hereinbefore, the entire 18.2 kb sequence of spanning Exons I through Exon X (referred to as 18-5665) inserted into bacteriophage lambda DNA was deposited under the accession number ATCC 40856. Exons XI and XII may be made from the sequence of Figure 1 using known techniques or isolated from the various cell lines noted above in which MSF cDNA has been detected.

The MSF DNA obtained as described above or modified as described above may be introduced into a selected expression vector to make a recombinant molecule or vector for use in the method of expressing novel MSF polypeptides. These vectors contain the novel MSF DNA sequences recited herein, which alone or in combination with other sequences, code for MSF polypeptides of the invention or active fragments thereof. The vector employed in the method also contains selected regulatory sequences in operative association with the DNA coding sequences of the invention. Regulatory sequences preferably present in the selected vector include promoter fragments, terminator fragments and other suitable sequences which direct the expression of the protein in an appropriate host cell. The resulting vector is capable of directing the replication and expression of an MSF in selected host cells. The transformation of these vectors into appropriate host cells can result in expression of the MSF polypeptides.

Numerous types of appropriate expression vectors are known in the art for mammalian (including human) expression, as well as insect, yeast, fungal and bacterial expression, by standard molecular biology techniques. Mammalian cell expression vectors are desirable for expression. Bacterial cells, e.g., E. coli, are also desirable for expression of MSFs.

Mammalian cell expression vectors described herein may be synthesized by techniques well known to those skilled in this art. The components of the vectors, e.g. replicons, selection genes, enhancers, promoters, and the like, may be obtained from natural sources or synthesized by known procedures. See, Kaufman et al, J. Mol. Biol. 159:511-521 (1982); and Kaufman, Proc. Natl. Acad. Sci. 82:689-693 (1985). Alternatively, the vector DNA may include all or part of the bovine papilloma virus genome [Lusky et al, Cell 36:391-401 (1984)] and be carried in cell lines such as C127 mouse cells as a stable episomal element.

One such vector for mammalian cells is pXM [Yang, Y. C. et al, Cell 47:3-10 (1986)]. This vector contains the SV40 origin of replication and enhancer, the adenovirus major late promoter, a cDNA copy of the adenovirus tripartite leader sequence, a small hybrid intervening sequence, an SV40 polyadenylation signal and the adenovirus VA I gene, in appropriate relationships to direct the high level expression of the desired cDNA in mammalian cells [See, e.g., Kaufman, Proc. Natl. Acad. Sci. 82:689-693 (1985)]. To generate constructs for expression of MSF, the pXM vector is linearized with an appropriate restriction endonuclease enzyme and separately ligated to the cDNA encoding MSF which has been appropriately prepared by restriction endonuclease digestion, for example.

Another similar vector is pMT21. pMT21 is prepared by linearizing pMT2pc (which has been deposited with the ATCC under Accession No. 40348) by digestion with PstI. The DNA is then blunted using T_4 DNA polymerase. An oligonucleotide:

TGCAGGCGAG CCTGAATTCC TCGA 24

is then ligated into the DNA, recreating the PstI site at the 5' end and adding an EcoRI site and XhoI site before the ATG of the DHFR cDNA. This plasmid is called pMT21. Preferably a desired polylinker with restriction sites for NotI, KpnI, SalI and SnabI is introduced into this vector for ready insertion of the MSF coding sequence.

Still another vector which may be employed to express MSF in CHO cells is pED4DPC-1. This vector is prepared from pED4, also known as pEMC2B1. As does pXM, described above, this vector contains the SV40 origin of replication and enhancer, the adenovirus major late promoter, a cDNA copy of the majority of the adenovirus tripartite leader sequence, a small hybrid intervening sequence, an SV40 polyadenylation signal and the adenovirus VA I gene, in appropriate relationships to direct the high level expression of the desired cDNA in mammalian cells. In addition, it contains DHFR and β-lactamase markers and an EMC sequence which pXM does not contain. To made pED4DPC-1, the sequence 1075 through 1096 is removed from pED4 to remove a stretch of cytosines. A new polylinker is added to introduce the restriction sites NotI, SalI and SnabI to the plasmid. The vector is linearized with an appropriate endonuclease enzyme and subsequently ligated separately to the cDNA encoding MSF.

These above-described vectors do not limit this invention, because one skilled in the art can also construct other useful mammalian expression vectors by, e.g., inserting the DNA sequence of the MSF from the plasmid with appropriate enzymes and employing well-known recombinant genetic engineering techniques and other known vectors, such as pJL3 and pJL4 [Gough et al., EMBO J. 4:645-653 (1985)] and pMT2 (starting with pMT2-VWF, ATCC No. 67122; see PCT Publication WO87/04187).

Once the vector is prepared, a selected host cell is transformed by conventional techniques with the vector containing MSF. The method of this present invention therefore comprises culturing a suitable cell or cell line, which has been transformed with a DNA sequence coding for expression of an MSF

polypeptide under the control of known regulatory sequences.

Suitable cells or cell lines may be mammalian cells, such as Chinese hamster ovary cells (CHO) or the monkey COS-1 cell line. CHO cells are preferred as a mammalian host cell of choice for stable integration of the vector DNA, and for subsequent amplification of the integrated vector DNA, both by conventional methods. The selection of suitable mammalian host cells and methods for transformation, culture, amplification, screening and product production and purification are known in the art. See, e.g., Gething and Sambrook, Nature 293:620-625 (1981), or alternatively, Kaufman et al, Mol. Cell. Biol. 5(7):1750-1759 (1985) or Howley et al, U. S. Patent No. 4,419,446. Another suitable mammalian cell line is the CV-1 cell line. Further exemplary mammalian host cells include particularly primate cell lines and rodent cell lines, including transformed cell lines. Normal diploid cells, cell strains derived from in vitro culture of primary tissue, as well as primary explants, are also suitable. Candidate cells may be genotypically deficient in the selection gene, or may contain a dominantly acting selection gene. Other suitable mammalian cell lines include, but are not limited to, HeLa, mouse L-929 cells, 3T3 or 293 lines derived from Swiss, Balb-c or NIH mice, BHK or HaK hamster cell lines.

Similarly useful as host cells suitable for the present invention are bacterial cells. For example, the various strains of E. coli (e.g., HB101, MC1061 and strains used in the following examples) are well-known as host cells in the field of biotechnology. When used as host cells, E. coli permits the expression of the MSF protein as a single protein. MSF may also be expressed in bacterial cells as a fusion protein with thioredoxin, as disclosed in detail in United States Patent Application Serial No. 07/652,531, which is incorporated herein by reference. Various strains of B. subtilis, Pseudomonas, other bacilli and the like may also be employed in this method.

Many strains of yeast cells known to those skilled in the

art are also available as host cells for expression of the polypeptides of the present invention. Additionally, where desired, insect cells may be utilized as host cells in the method of the present invention. See, e.g. Miller et al, Genetic Engineering 8:277-298 (Plenum Press 1986) and references cited therein. Fungal cells may also be employed as expression systems.

Once the MSF is expressed by the transformed and cultured cells, it is then recovered, isolated and purified from the culture medium (or from the cell, if expressed intracellularly) by appropriate means known to one of skill in the art.

A preferred purification procedure to isolate a recombinant or synthetic MSF from serum free mammalian cell (COS-1) conditioned medium is characterized by steps are similar to those for the purification of native Meg-CSF from urine and are described in detail in Example 4. The recombinant MSF is concentrated from COS-1 cell supernatant through an Amicon YM-10 filter with a 10,000 Dalton molecular weight cut-off. The concentrate is dialyzed into 20mm sodium acetate, pH4.5, and then applied to an S Toyopearl cation exchange FPLC column equilibrated in 20mM sodium acetate, pH4.5. The bound material is then eluted from the column acidified with 10% TFA to 0.1% TFA and applied through a cycle of C4 reverse phase HPLC using 0.1% TFA/acetonitrile as the solvent system. In the case of MSF-K130, the protein elutes between 20-35% of a buffer containing 0.1% TFA, 95% acetonitrile. Other non-naturally occurring MSFs described above may be obtained by applying this purification scheme described in detail in Example 4 for MSF-K130.

MSF polypeptides may also be produced by known conventional chemical synthesis, e.g., by Merrifield synthesis or modifications thereof. Methods for constructing the polypeptides of the present invention by synthetic means are known to those of skill in the art. The synthetically-constructed MSF polypeptide sequences, by virtue of sharing

primary, secondary, or tertiary structural and conformational characteristics with native MSF polypeptides may possess MSF biological properties in common therewith. Thus, they may be employed as biologically active or immunological substitutes for natural, purified MSF polypeptides in therapeutic and immunological processes.

One or more MSFs or active fragments thereof, purified in a homogeneous form or as a mixture of different MSFs from natural sources or produced recombinantly or synthetically, may be used in a pharmaceutical preparation or formulation. The MSF pharmaceutical compositions of the present invention or pharmaceutically effective fragments thereof may be employed in the treatment of immune deficiencies or disorders. MSFs may also be employed to treat deficiencies in hematopoietic progenitor or stem cells, or disorders relating thereto. MSFs may be employed in methods for treating cancer and other pathological states resulting from disease, exposure to radiation or drugs, and including for example, leukopenia, bacterial and viral infections, anemia, B cell or T cell deficiencies, including immune cell or hematopoietic cell deficiency following a bone marrow transplantation. MSFs may also be used to potentiate the immune response to a variety of vaccines creating longer lasting and more effective immunity. MSFs may be employed to stimulate development of B cells, as well as megakaryocytes.

The MSFs of the present invention may also have utility in stimulating platelet production, stimulating platelet recovery following chemotherapy or bone marrow transplantation, treating thrombocytopenia, aplastic anemia and other platelet disorders, preserving and extending the lifetime of platelets in storage, and stimulating platelet production <u>in vitro</u> for use in platelet transfusions. MSFs may also be employed to stimulate the growth and development of other colonies of hematopoietic and non-hematopoietic cells. Similarly, these factors may be useful in cell-targeting. MSF may also be useful in the treatment of

wounds or burns, alone or with other wound-healing agents, such as fibroblast growth factor (FGF). MSFs are believed to have adhesion molecule type properties and thus, known therapeutic uses of such adhesion molecules are also contemplated for MSFs of this invention. MSF compositions may be used as an adjunctive therapy for bone marrow transplant patients.

Therapeutic treatment of such platelet disorders or deficiencies with these MSF polypeptide compositions may avoid undesirable side effects caused by treatment with presently available serum-derived factors or transfusions of human platelets. It may also be possible to employ one or more active peptide fragments of MSF in such pharmaceutical formulations.

Therefore, as yet another aspect of the invention are therapeutic compositions for treating the conditions referred to above. Such compositions comprise a therapeutically effective amount of a MSF protein, a therapeutically effective fragment thereof, or a mixture of variously spliced or otherwise modified MSFs in admixture with a pharmaceutically acceptable carrier. This composition can be systemically administered parenterally. Alternatively, the composition may be administered intravenously. If desired, the composition may be administered subcutaneously. When systemically administered, the therapeutic composition for use in this invention is in the form of a pyrogen-free, parenterally acceptable aqueous solution. The preparation of such pharmaceutically acceptable protein solutions, having due regard to pH, isotonicity, stability and the like, is within the skill of the art.

The dosage regimen involved in a method for treating the above-described conditions will be determined by the attending physician considering various factors which modify the action of drugs, e.g. the condition, body weight, sex and diet of the patient, the severity of any infection, time of administration and other clinical factors. Generally, the daily regimen should be in the range of about 1 to about 1000 micrograms of MSF protein, mixture of MSF proteins or fragments thereof.

Alternatively about 50 to about 50,000 units (i.e., one unit being the minimum concentration of MSF protein, or MSF protein mixture, which yields the maximal number of colonies in the murine fibrin clot megakaryocyte colony formation assay) of MSF protein per kilogram of body weight may be a desirable dosage range.

The therapeutic method, compositions, purified proteins and polypeptides of the present invention may also be employed, alone or in combination with other cytokines, hematopoietins, interleukins, growth factors or antibodies in the treatment of disease states characterized by other symptoms as well as platelet deficiencies. It is anticipated that an MSF, if it does not itself have TPO activity, will prove useful in treating some forms of thrombocytopenia in combination with general stimulators of hematopoiesis, such as IL-3, IL-6, GM-CSF, steel factor, IL-11 (described in PCT Publication WO91/07495) or with other megakaryocytic stimulatory factors or molecules with TPO-like activity. Additional exemplary cytokines or hematopoietins for such co-administration include TPO, G-CSF, the M-CSFs, IL-1, IL-4, IL-7, erythropoietin, and variants of all of these cytokines, or a combination of multiple cytokines. The dosage recited above would be adjusted to compensate for such additional components in the therapeutic composition. For example, the MSF may be administered in amounts from 1 to 1000 μg/kg body weight and the other cytokine may be administered in the same amounts in such a co-administration protocol. Alternatively, co-administration may permit lesser amounts of each therapeutic agent to be administered. Progress of the treated patient can be monitored by conventional methods.

Other uses for these novel proteins and recombinant polypeptides are in the development of antibodies generated by standard methods for <u>in vivo</u> or <u>in vitro</u> diagnostic or therapeutic use. As diagnostic or research reagents, antibodies generated against these MSFs may also be useful in affinity columns and the like to further purify and identify the complete

Meg-CSF protein. Such antibodies may include both monoclonal and polyclonal antibodies, as well as chimeric antibodies or "recombinant" antibodies generated by known techniques.

The antibodies of the present invention may be utilized for in vivo and in vitro diagnostic purposes, such as by associating the antibodies with detectable labels or label systems. Alternatively these antibodies may be employed for in vivo and in vitro therapeutic purposes, such as by association with certain toxic or therapeutic compounds or moieties known to those of skill in this art. These antibodies also have utility as research reagents.

Also provided by this invention are the cell lines generated by presenting MSF or a fragment thereof as an antigen to a selected mammal, followed by fusing cells of the animal with certain cancer cells to create immortalized cell lines by known techniques. The methods employed to generate such cell lines and antibodies directed against all or portions of a human MSF protein or recombinant polypeptide of the present invention are also encompassed by this invention.

Examples

The following examples illustratively describe the purification and characteristics of homogeneous human MSF and other methods and products of the present invention. These examples are for illustration and are not intended to limit the scope of the present invention.

Example 1 - Purification and Biochemical Characteristics of Native Meg-CSF from Urine

The following procedures are employed to obtain native Meg-CSF protein from urine of human bone marrow transplant patients. The same or similar procedure may be employed to purify other MSFs from natural sources. Urine from patients with aplastic anemia or thrombocytopenia accompanying other disease states may also be used as the source of the factor employing this

purification.

STEP 1: Urine was collected from bone marrow transplant patients between days 5 and 18 following transplant. Between fifty and one hundred liters of pooled urine were treated with protease inhibitors phenylmethyl-sulfonylfluoride (PMSF) and ethylenediaminetetraacetic acid (EDTA). This pooled urine was concentrated on an Amicon YM-10 filter (10,000 molecular weight cut-off) to remove excess pigments and reduce the volume. A cocktail of protease inhibitors [leupeptin, pepstatin, ethylene glycol-bis-tetraacetic acid (EGTA) and N-ethylmaleimide (NEM)] was added to the urine at this step and the next three steps to minimize proteolysis. The pH of the urine concentrate was adjusted to 8.0 and diluted to a conductivity of 7mS/cm.

STEP 2: The retentate from this first step of the purification was then subjected to anion exchange column chromatography on a QAE Zetaprep [Cuno] at pH 8.0. The QAE flow-through was adjusted to a pH4.5 with 1M acetic acid.

STEP 3: The flow-through from the second purification step was bound to a cation exchange chromatographic column, an SP-Zetaprep column [Cuno] at pH 4.5. Bound protein containing Meg-CSF was eluted with 1M NaCl at a pH of 4.5. The eluate was pooled, protease inhibitors were added as above and the bound protein was either neutralized to pH 7 and stored at -80°C until further chromatography was performed or dialyzed into a Tris-buffered saline (TBS) solution, with the addition of the protease inhibitors described in Step 1. This dialyzate was heated at 56°C for 30 minutes. Addition of the protease inhibitors, while not essential for recovery of protein, enabled greater amount of protein to be recovered from this step, by inactivating the proteases in the system. Pools from this step were also analyzed for the presence of megakaryocyte-specific growth factors. These pools were found to contain Meg-CSF activity.

STEP 4: The resulting material was added to a lectin affinity chromatographic column, a Wheat Germ Sepharose column

[Pharmacia]. Urinary Meg-CSF was found to bind to this column. This protein was then eluted with 0.25 M N-acetyl glucosamine (N-acglcNH_2) in TBS, and dialyzed against 20 mM sodium acetate, pH 4.5 in the presence of the protease inhibitors of Step 1.

STEP 5: This dialysate was applied to a 10 ml S-Toyopearl FPLC cation exchange column and eluted using a linear gradient of 0 to 1M NaCl in 20mM sodium acetate at pH 4.5. The protein eluted from this step was tested for Meg-CSF activity in the fibrin clot assay described below. The Meg-CSF activity was observed to elute in two discrete peaks. The major activity eluted between 0.1M and 0.25M NaCl. A minor, but reproducible activity eluted between 0.3M and 0.5N NaCl. The two activities may be due to protein or carbohydrate modification of a single protein; however the data presented further herein refers to the major protein.

STEP 6: The eluate from this fifth purification step was then purified on a reverse phase HPLC (C4) column [Vydac; 1cmX25cm] using 0.1% TFA (trifluoroacetic acid) in 95% acetronitrile. This step removes an abundant 30Kd protein contaminant. Recombinant MSF elutes at a slightly lower gradient of about 20 - 33% of the B buffer.

STEP 7: The HPLC step was repeated in a different solvent system, after the eluate of Step 6 was diluted with two parts acetic acid and pyridine. The purified material eluted between 6-15% n-propanol in pyridine and acetic acid on a C18 reverse phase HPLC column (0.46 X 25 cm). The material produced after this step, when assayed, gave a specific activity of greater than 5 X 10^7 dilution units per milligram in the murine assay of Example 10. This optional step removes the bulk of urinary ribonuclease, a major contaminant, from the preparation.

STEP 8: The HPLC step was repeated once more on a C4 column (Vydac; 0.46 X 25 cm) using 0.15% HFBA in acetonitrile. The material eluted between 27-37% acetonitrile. The last HPLC step removed substantially all remaining ribonuclease and proteinaceous contaminants present after Step 7.

This purified Meg-CSF material was then analyzed by SDS-PAGE, bioassayed and labelled with ^{125}I. Homogenous protein is obtained from this procedure, omitting step 7, having a specific activity ranging from about $5X10^7$ to about $2\text{-}5X10^8$ dilution units per mg protein in the murine megakaryocyte colony assay described in Example 8. A unit of activity is defined as the reciprocal of the maximum dilution which stimulates the maximum number of megakaryocyte colonies per ml.

This process is preferably used to purify any naturally occurring MSF protein, or mixture thereof, from a natural source.

Other physical and chemical properties of this urinary Meg-CSF were determined as follows:

The molecular weight of the protein was found to be about 35-45 kD as measured by SDS-PAGE on 12% gels run at 60 mA for 2 hours under non-reducing conditions. Under reducing conditions [10mM DTT (dithiothreitol)] on 12% SDS-PAGE, the molecular weight was 22-25 kD. Urinary Meg-CSF appears to be a homodimer.

The eluate from step 2 of the above purification was dialyzed in tris buffered saline (TBS) at pH 7.5 overnight, then loaded onto a 200 ml wheat germ sepharose column overnight at 0.5 col. vol./hr. After loading, the column was washed with TBS to remove unbound proteins until A^{280} baseline was reached. The urinary Meg-CSF protein bound to the wheat germ sepharose and was eluted in TBS and 0.25M N-acetylglucosamine, indicating that the protein is glycosylated.

Upon N-glycanase digestion under non-reducing conditions in 1mg/ml BSA, 1.7% SDS, 0.2M $NaHPO_4$ at pH 8.4 and 5mM EDTA, the urinary protein was determined to contain no N-linked carbohydrate.

Presumably, the protein contains O-linked carbohydrate. The urinary Meg-CSF protein failed to bind heparin sepharose when a dialyzed sample of the eluate from Step 6 above was loaded onto the column in 20mM tris-Cl at pH 7.4 and eluted with 20mM tris-Cl and 1M NaCl (pH 7.4).

When run on reverse phase HPLC with a C4 vydac column using an A buffer of 0.1 % TFA, a B buffer of 95% acetonitrile in 0.1% TFA, and a gradient from 5-100% using buffer B, urinary Meg-CSF elutes between 20-35% acetonitrite.

Example 2 - Analysis of Genomic MSF, Meg-CSF

A preliminary analysis of COS supernatant expressing the Kpn-SnaBl 12 kb genomic subclone isolated as described in WO91/02001 was performed and indicated that a protein which reacted with MSF-specific antibodies was expressed by COS cells and was secreted into the culture medium. Dialyzed, concentrated cell supernatant was variably active in the murine meg-colony assay.

Analysis by Northern and Western indicated that the level of MSF mRNA and protein was very low. A western immunoblot of the protein from COS supernatant expressed in conditioned medium revealed the presence of three heterogenous species which were shown to specifically bind anti Meg-CSF peptide antibodies by competition for the antibodies with excess peptide. The molecular weights of these species, 200 kD, 30 kD, and 14 kD, are different from the partially purified Meg-CSF from the BMT urine which has an apparent molecular weight ranging between about 16 to 21 kD on 10-20% gradient SDS-PAGE run at 60 mA for one hour under reducing conditions (10mM DDT).

Example 3 - Construction and Mammalian Cell Expression of Recombinant MSFs

Twelve MSF cDNA clones, truncated by known techniques, were constructed by using the polymerase chain reaction. A thirteenth clone, MSF-K130, was isolated as a consequence during the PCR reaction by the inadvertent insertion of an artificial termination codon in Exon IV after amino acid 130. Thirteen oligonucleotide primers were synthesized as follows:

(1) CGCGCGGCCGCGACTATTCG

(2) GCGCTCGAGCTAAGAGGAGGAGGA

(3) GCGCTCGAGCTATCTATTAGCAGC
(4) GCGCTCGAGCTACTTGTTATCTTT
(5) GCGCTCGAGCTAATCTACAACTGG
(6) GCGCTCGAGCTAGTTTGGTGGTTT
(7) GCGCTCGAGCTAAGTTCTGTTCTT
(8) GCGCTCGAGCTAGGTTGTTGATTT
(9) GCGCTCGAGCTAACGTTTGGTTGT
(10) GCGCTCGAGCTATGGTTTGGGTGA
(11) GCGCTCGAGCTACTTCTTCTTGTT
(12) GCGCTCGAGCTATTTCTTAGTCTT
(13) GCGCTCGAGCTATTCTTCTGTTAT

Primer (1) was designed to hybridize to the cDNA flanking the initiating methionine codon. It contains nine MSF-homologous nucleotides, a NotI restriction endonuclease site and three additional nucleotides to enhance restriction endonuclease recognition (as suggested in the New England Biolabs catalog).

Oligonucleotide primers (2) through (7) were designed to hybridize to the 3' regions of the cDNA and to flank the putative MSF protein processing site codons for S172 [primer (2) above], MSF-R192 [primer (3) above], MSF-K204 [primer (4) above], MSF-K130 and D220 [primer (5) above], N141 [primer (6) above] and T208 [primer (7) above]. These 3' primers contain twelve nucleotides of MSF-homologous sequence, a translation termination codon, an XhoI restriction endonuclease site and three additional nucleotides to enhance restriction endonuclease recognition.

Six PCR reactions were performed in duplicate, using the conditions recommended by Perkin-Elmer Cetus Corp. Each of the six duplicate reactions contained the 5' primer (No. 1, 1.0 μM), one of the 3' primers (1.0 μM) and 1 ng of MSF cDNA as template. The reactions were carried through two rounds of eighteen cycles each. One cycle consisted of a two minute denaturation of 95°C followed by three minutes of annealing/extension of 72°C. After the first round of eighteen cycles, 10 μl of the first reaction was transferred to a fresh reaction mixture (100 μl total), and

the amplification cycles were repeated. The PCR products generated by the second round of amplification reactions (twelve in all) were digested with NotI and XhoI, using conditions described by the vendor, and fractionated by agarose gel electrophoresis.

To obtain expression of these truncated MSFs in mammalian host cells, the appropriate DNA bands were then excised and ligated into a NotI and XhoI digested pMT21-2 vector. The vector pMT21-2 is identical to the vector pMT21 except for the polylinker region, containing PstI, NotI, KpnI, ApaI, EcoRV, EcoRI and XhoI sites, which was changed to facilitate cloning of MSF DNA fragments. Competent DH5 cells were transformed with the recombinant plasmid and selected for resistance to ampicillin. Plasmid DNA was prepared from transformed cells and sequenced with selected internal oligonucleotides across the MSF insert. All the above techniques are conventional and described in Sambrook et al, cited above.

The clones listed above were identified as having the correct nucleotide sequence to encode the desired MSF proteins. For example, S172 encodes MSF amino acids 1-172, terminating with a serine residue. Position 173 encodes a translation termination codon. The exception was one of the two reactions designed to synthesize D220. During this PCR reaction, a serendipitous deletion of nucleotide 392 of the cDNA sequence resulted in clone MSF-K130, which encodes MSF amino acids 1-130 and terminates in a lysine followed by a TAA stop codon. Clone MSF-K130 may readily be deliberately synthesized by a PCR reaction designed for this purpose. This would require using an oligonucleotide primer similar in design to the other 3' primer oligonucleotides, i.e., an oligonucleotide containing twelve nucleotides of MSF-homologous sequence, a translation termination codon, an XhoI restriction site and a few additional nucleotides to enhance restriction endonuclease recognition. An example of a suitable 3' primer would be the following sequence: GCGCTCGAGCTAATTTGATGGTTT.

Six additional mutants were synthesized with some modification in the procedure described above. Primer (1) was used as the 5' primer in reactions with oligonucleotide primers designed to hybridize to the 3' regions of the cDNA flanking the putative MSF protein processing site codons for MSF-T133 [primer (8) above], MSF-R135 [primer (9) above], MSF-P139 [primer (10) above], MSF-K144 [primer (11) above], MSF-K147 [primer (12) above], and MSF-E157 [primer (13) above].

The PCR reactions were performed in duplicate and contained the 5' primer (No. 1, 1.0 μM), one of the 3' primers (1.0 μM) and 1 μg of MSF-R192 (from the first set of reactions) as template. The reactions were carried through twenty five cycles consisting of 2 rounds of 18.

The PCR products were digested with NotI and XhoI, using conditions described by the vendor and fractionated by agarose gel electrophoresis. The appropriate DNA bands were then excised and ligated into pED4DPC-1 NOT/SAL, a PMT21 derivative.

Expression is accomplished as follows: The PMT21-2 plasmid, containing the MSF DNA sequence is transfected onto COS cells. The conditioned medium from the transfected COS cells contains MSF biological activity as measured in the murine assays. Similarly the modified pED4DPC-1 construct containing the cDNA for MSF is transfected into CHO cells.

The vector pED4DPC-1 may be derived from pMT21 vector. pMT21 is cut with EcoRI and XhoI which cleaves the plasmid at two adjacent cloning sites. An EMCV fragment of 508 base pairs is cut from pMT_2ECAT_1 [Jong, S. K. et al, J. Virol. 63:1651-1660 (1989)] with the restriction enzymes EcoRI and TaqI. A pair of oligonucleotides 68 nucleotides in length are synthesized to duplicate the EMCV sequence up to the ATG. The ATG is changed to an ATT, and a C is added, creating a XhoI site at the 3' end. A TaqI site is situated at the 5' end. The sequences of the oligonucleotides are:

CGAGGTTAAA AAACGTCTAG GCCCCCCGAA CCACGGGGAC 40

GTGGTTTTCC TTTGAAAAAC ACGATTGC 68

and the respective complementary strands.

Ligation of the pMT21 EcoRI-to-XhoI fragment to the EMCV EcoRI-to-TaqI fragment and to the TaqI/XhoI oligonucleotides produces the vector pED4. A polylinker, containing PstI, NotI, SalI, SnaBI and EcoRI sites, is inserted into this vector as described above to create pED4PC-1.

Stable transformants are then screened for expression of the product by standard immunological, biological or enzymatic assays, such as those described below in Example 8. The presence of the DNA and mRNA encoding the MSF polypeptides is detected by standard procedures such as Southern and Northern blotting. Transient expression of the DNA encoding the polypeptides during the several days after introduction of the expression vector DNA into suitable host cells is measured without selection by activity or immunologic assay, e.g., the murine fibrin clot assay, of the proteins in the culture medium.

Example 4 - Purification and Biochemical Characteristics of MSF-K130 from COS cells.

An initial 3L batch of serum-free conditioned medium from COS-1 cells transfected with MSF-K130 yielded 140 ug of purified, active MSF protein using a three step purification process. COS-1 cell conditioned medium harvested under serum free conditions was concentrated on an Amicon YM10 membrane with a molecular weight cutoff of 10,000 daltons. The concentrate was centrifuged at 10,000 rpm in an SS34 rotor at 4°C to remove cellular debris and precipitate. The supernatant was dialyzed against 20mM sodium acetate pH 4.5 overnight at 4°C. The dialyzed protein solution was centrifuged again at low speed to remove residual precipitate. The dialyzed MSF-K130 containing solution was applied to an S Toyopearl cation exchange FPLC column, equilibrated in 20mM sodium acetate pH 4.5. Bound protein was eluted with a gradient of 0 to 1M NaCl in 20mM sodium acetate, pH 4.5, at room temperature. The protein that

eluted from this step was tested for Meg-CSF activity in the fibrin clot assay described below.

MSF-K130 eluted between 0.1 to 0.2 M NaCl. The active MSF peak was observed to have a molecular weight on SDS-PAGE 10-20% gradient polyacrylamide gels of between 20-45 kD under non-reducing conditions and 18-21 kD under 10mM DTT reducing conditions. The molecular weight of MSF-K130 S toyo FPLC fractions was determined by a western immunoblot of S toyo pearl fractions using anti Meg-CSF peptide rabbit antisera.

The pool of active MSF-K130 was divided into three aliquots based on the molecular weights under non-reducing conditions. Pool A consisted of mostly high molecular weight dimer, 35-45 kD. Pool B consisted of intermediate molecular weight dimer species ranging from 20-45 kD; and pool C comprised predominantly monomer species of molecular weight range 14-25 kD. MSF from all three pools had a molecular weight of 18-21 kD under reducing conditions.

The final purification step was one cycle of reverse phase-HPLC. Protein from the three pools were acidified with 10% TFA to 0.1% TFA (v/v), filtered through a 0.45 μm PVDF membrane and injected at 1ml/min onto a 25cm x 4.6mm C4 (Vydac) reverse phase HPLC column equilibrated in 0.1% TFA at room temperature. Bound protein was eluted with a linear gradient of 0-95% acetonitrile in 0.1% TFA. Typically, Meg-CSF activity eluted between 15-30% buffer B (95% acetonitrile in 0.1% TFA).

The molecular weight of MSF-K130 ranged between 20-45 kD as measured by SDS-PAGE in 10-20% gels run at 60 mA for 1 hour under non-reducing conditions. Under reducing condition of 10mM dithiothreitol (DTT), the material yields a molecular weight of between 18-21 kD, a major species was detected of about 19 kD.

Upon N-glycanase digestion under non-reducing or 10mM DTT reducing conditions in 1 mg/ml BSA, 1.7% SDS, 0.2 M $NaHPO_4$ at pH 8.4 and 5mM EDTA, the MSF-K130 protein was determined to contain no N-linked carbohydrate. MSF-K130 protein bound Jacalin C agarose, an O-linked carbohydrate binding lectin, indicating the

presence of O-linked carbohydrates. Specifically, serum-free conditioned media from COS-1 cells transfected with MSF-K130 was dialyzed in 0.175 M tris-Cl, 0.15M NaCl 0.1 mM $CaCl_2$ pH 7.4 buffer. Two mls of the dialyzed material was added to 1 ml Jacalin C agarose equilibrated in the same buffer and allowed to bind the resin overnight at 4°C. The protein solution that did not bind Jacalin C agarose was collected and the resin was washed with 10 column volumes of the starting buffer. Protein bound to the Jacalin C agarose resin was eluted with the starting buffer plus 20mM α methyl galactopyranoside. MSF-K130 was detected in the Jacalin C eluate but not the Jacalin C unbound by Western immunoblot using anti Meg-CSF peptide rabbit antisera.

MSF-K130 protein did not bind to heparin agarose under standard binding conditions. Specifically 2 mls of serum-free conditioned medium from COS-1 cells transfected with MSF-K130 was dialyzed into 20mM tris-Cl, pH 7.4 buffer. The dialyzed protein solution was loaded onto a 0.2ml heparin agarose column equilibrated in the same buffer and allowed to bind the resin overnight at 4°C. The protein solution that did not bind heparin agarose was collected and the resin was washed with 10 column volumes of starting buffer. Protein that bound to the heparin agarose was eluted with 20 mM tris-Cl, 1M NaCl, pH 7.4. MSF-K130 was detected in the heparin unbound but not the heparin bound by western immunoblot using anti Meg-CSF peptide rabbit antisera.

When run on reverse phase HPLC with a C4 Vydac column using an A buffer of 0.1% TFA, B buffer of 95% acetonitrile in 0.1% TFA, and a linear gradient, MSF-K130 eluted between 20-35% acetonitrile. The theoretical isoelectric point was calculated to be 5.76.

The specific activity of MSF-K130 ranged from 1.9×10^7 to 8.6×10^7 dilution units per mg protein in the murine megakaryocyte colony assay described in Example 10.

Example 5.- Isolation of a 4Aλ cDNA Clone

A recombinant 4Aλ cDNA clone (corresponding to amino acids 1 through 924 of Figure 1 and therefore also termed MSF-L924) may be isolated by standard molecular biology techniques. COS cells are transfected with the Kpn-SnaBl genomic subclone, and from these cells polyA$^+$ RNA is isolated. A cDNA library is prepared from this RNA by subcloning EcoRI adapted cDNA into the cloning vector LAMBDA ZAP available from Stratagene Inc. Recombinant phage composing the library are plated and duplicate nitrocellulose and/or nylon replicas are made of the plates.

Clones containing Exons I, II, III, IV, V and VI are identified by probing the replicas with ^{32}p labeled oligonucleotide probes. The probes are 24mers designed to span the junctions between the successive exons, for example a probe with 12 nucleotides complementary to Exon I and 12 nucleotides complementary the immediately adjacent nucleotides in Exon II. Following hybridization and autoradiography, the filters can be stripped of radioactivity with 0.1N NaOH and reprobed with an oligonucleotide which spans the junction between Exons II and III. Phage which hybridize to both probes are identified, replated and probed with oligonucleotides which span exon junctions III/IV and IV/V. Phage which hybridize to both probes are identified, replated and probed with an oligonucleotide which spans the V/VI junction. Because of the low frequency of clones which contain Exon V, it is easiest if clones which contain this exon are identified first.

MSF inserts within the chosen phage clones are excised by using the Automatic Excision Protocol as described by and available from Stratagene Inc. This yields subclones in the vector Pbluescript SK-. These clones are characterized by restriction mapping and DNA sequencing and verified by comparison to the MSF cDNA sequence of Figure 1. One clone that was isolated (termed clone "21A") contained all of Exons I, II, III, IV and V but did not contain all of Exon 6 to the SnaBI site.

A clone containing all of Exons I, II, III, IV, V and to the SnaBI site in Exon VI was prepared as follows. Clone 21A was digested with the enzymes AccI and NotI. The human genomic clone 18-5665 was digested with AccI and SnaBI. The appropriate MSF DNA fragments from both digests were isolated from agarose gels following electrophoresis and ligated together. The ligated products were digested with the enzymes NotI and SnaBI, again separated by electrophoresis, and the band corresponding to the NotI-AccI::AccI-SnaBI ligation product was purified. This fragment was subcloned into NotI and EcoRV digested pMT21-2 and transformed into bacterial cells as described above to yield the clone 4A1.

The 4A1 clone contains at its 3' end some additional amino acids which are encoded by the vector. The last few amino acids in this clone are KTTERDLHHLPEFLEPSWFDH*. The underlined amino acids are not found in the MSF gene; the * denotes a termination codon. Clones which do not contain these additional non-MSF amino acids may be constructed by ligating the NotI-AccI::AccI-SnaBI fragment described above into a pMT21 derivative which would contain a SnaBI site in the polylinker followed by an in-frame translation termination codon. An example of such a sequence would be TACGTACATAA. The SnaBI site is TACGTA, and the TAA encodes an in-frame termination codon so that the protein produced will contain only MSF amino acids.

Alternatively, if clones containing sequence to the Exon VI SnaBI site are obtained, they can be expressed directly by digestion with the restriction enzymes NotI and XhoI. The MSF insert is separated from the vector Pbluescript by electrophoresis into agarose gels, excised and ligated into a NotI and XhoI digested expression vector pMT21-2. Competent bacterial cells are transformed and selected for resistance to ampicillin. Plasmid DNA prepared from the bacteria is transfected into COS cells using standard techniques.

Example 6 - Purification and Biochemical Characteristics of (4aλ) from COS Cells

Recombinant MSF 4aλ (MSF-L924) was purified from 5L COS-1 cell conditioned medium using a 3 step purification process. COS-1 cell conditioned medium harvested under serum-free conditions was concentrated on a YM10 Amicon membrane with a molecular cutoff of 10,000 daltons. The concentrate was centrifuged at 10,000 rpm in an SS34 rotor at 4°C to remove cellular debris and precipitate.

The supernatant was dialyzed against 20mM sodium acetate pH 5.0, .02% tween 20 overnight at 4°C. The dialyzed protein solution was centrifuged again at low speed to remove residual precipitate. The dialyzed MSF 4aλ-containing solution was applied to a PEI anion exchange column available from J. T. Baker Company, equilibrated in 20 mM sodium acetate, 0.2% tween 20, pH 5.0, at 4°C. The column was then washed with 10 column volumes of equilibration buffer and MSF 4aλ was eluted with 1 and 2M NaCl in the equilibration buffer. Recombinant MSF 4aλ was detected in the purification fractions both by western immunoblot using anti Meg-CSF peptide rabbit antisera and activity on a murine bone marrow megakaryocyte colony forming assay.

The PEI elution containing recombinant MSF 4aλ was dialyzed into TBS overnight at 4°C. The dialyzed PEI fraction containing recombinant MSF 4aλ was applied to a heparin Toyopearl FPLC column, equilibrated in TBS, pH 7.4. The resin was washed with 10 column volumes of TBS and recombinant MSF 4aλ was eluted with 0.3 and 0.5 M NaCl in a stepwise elution method at room temperature. The final purification step was 1 cycle of RP-HPLC. Protein from the 0.3 and 0.5 M NaCl heparin FPLC elution was acidified with 10% TFA to 0.1% TFA (v/v), filtered through a 0.45 micron PVDF membrane and injected at 1 ml/min onto a 25cm X 4.6mm C4 (Vydac) reverse phase column equilibrated in 0.1% TFA at room temperature. Bound protein was eluted with a linear gradient of 5-95% acetonitrile in 0.1% TFA. Typically,

recombinant MSF 4aλ activity eluted between 15-30% buffer B (95% acetonitrile in 0.1% TFA).

The molecular weight of the purified recombinant MSF 4aλ was determined by 4-20% SDS-PAGE and detected by both silver stain and western immunoblot. The major recombinant MSF 4aλ protein is greater than 200 kD and smaller forms are present with molecular weights ranging between 15 to 70 kD.

MSF 4aλ contains several different molecular weight protein forms from COS-1 cell conditioned media. These forms have molecular weights ranging between 15 to 200 kD non-reduced and reduced. The protein form with the lowest molecular weight does not bind heparin agarose under standard binding conditions, indicating that this protein does not have a heparin binding domain. All other protein forms do bind heparin agarose under standard binding conditions, indicating that these protein forms do contain heparin binding domains. Specifically, 30 mls of conditioned medium from COS-1 cells transfected with 4aλ was concentrated on a YM10 membrane with a molecular cutoff of 10,000 kD. The concentrate was centrifuged at 10,000 rpm in an SS34 rotor at 4°C overnight to remove cellular debris and precipitate. The concentrate was dialyzed into 20 mM tris-Cl pH 7.4 buffer. Two mls of the dialyzed material was added to 1 ml heparin agarose equilibrated in the same buffer and allowed to bind overnight at 4°C. The protein solution that did not bind heparin agarose was collected and the resin was washed with 10 column volumes of the starting buffer. Protein that bound to the heparin agarose was eluted with 20 mM tris-Cl 1M NaCl, pH 7.4. The 4aλ protein was detected in both the heparin unbound and heparin bound fractions by western immunoblot. The molecular weight of the protein in the heparin unbound fraction was 25 kD under non-reducing and 18 kD under reducing standard 10-20% gradient SDS-PAGE conditions. The molecular weight of the MSF heparin binding protein was between 20 to 200 kD under non-reducing and reducing standard 10-20% gradient SDS-PAGE conditions.

MSF 4aλ protein did not bind lentil lectin sepharose or Con A sepharose under standard binding conditions. Specifi-cally, 60 mls of conditioned medium from COS-1 cells transfected with 4aλ was concentrated to 4 mls on a YM10 membrane with a molecular cutoff of 10,000 kD. The concentrate was centrifuged at 10,000 rpm in an SS34 rotor at 4°C overnight to remove cellular debris and precipitate. One-half of the concentrate was dialyzed into TBS and the second half was dialyzed into TBA containing 1mM $MnCl_2$ 1mM $CaCl_2$. Two mls of the concentrated material dialyzed into TBS was loaded onto a lentil lectin sepharose column equilibrated into the same buffer and allowed to bind overnight at 4°C. The protein solution that did not bind lentil lectin sepharose was collected and the resin was washed with 10 column volumes of the starting buffer. Protein that bound to the lentil lectin sepharose was eluted with TBS containing 0.25M α methyl mannopyranoside. MSF 4aλ was detected in the lentil lectin sepharose unbound fraction but not in the lentil lectin bound fraction by western immunoblot using anti Meg-CSF peptide rabbit antisera. Similarly, two mls of the concentrated material dialyzed into TBS containing 1mM $MnCl_2$ 1mM $CaCl_2$ was loaded onto a Con A sepharose column equilibrated in the same buffer and allowed to bind overnight at 4°C. The protein solution that did not bind Con A sepharose was collected and the resin was washed with 10 column volumes of the starting buffer. Protein that bound to the Con A sepharose was eluted with TBS containing 1mM $MnCl_2$ 1mM $CaCl_2$ and 0.25M α methyl mannopyranoside. MSF was detected in the Con A unbound fraction but not in the Con A bound fraction by western immunoblot using anti Meg-CSF peptide rabbit antisera.

The theoretical isoelectric point of MSF 4aλ is 9.88.

When run on reverse phase HPLC on a C4 vydac column using an A buffer of 0.1% TFA, a B buffer of 95% acetonitrile in 0.1% TFA with a linear gradient, MSF 4aλ eluted between 15-30% acetonitrile. The specific activity of MSF 4aλ was between $1X10^6$ and $1X10^7$ dilution units per mg protein in the murine

megakaryocyte colony assay of Example 10.

Carbohydrate composition analysis was performed on 30 µg of the purified material by methanolysis, followed by derivatization of liberated monosaccharides to their trimethylsilyl ethers and subsequent gas chromatography following the procedures of Clamp, et al., Glycoproteins, their Composition, Structure and Function, Part A, Section 6, Ch. 3, Elsevier Publ. (1972) and Reinhold, Meth. Enzymol. 25:244-249 (1972). Based on an apparent molecular weight of 100 kD, the ratio of nanomole sugar per nanomole 4aA is:

Galactose	73
GalNAc	91
GlcNAc	9
Sialic Acid	14

Fucose and Mannose were not detectable in the sample analyzed.

NMR spectroscopy confirms the presence of this extensive posttranslational O-linked glycosylation.

Example 7 - Bacterial Expression Systems

One skilled in the art could manipulate the sequences encoding the MSF polypeptide by eliminating any human regulatory sequences flanking the coding sequences, eliminating the mammalian secretory sequence of Exon I, and inserting bacterial regulatory sequences to create bacterial vectors for intracellular or extracellular expression of the MSF polypeptide of the invention by bacterial cells. The DNA encoding the polypeptides may be further modified to contain different codons to optimize bacterial expression as is known in the art.

The sequences encoding the mature MSF are operatively linked in-frame to nucleotide sequences encoding a secretory leader polypeptide permitting bacterial expression, secretion and processing of the mature MSF polypeptides, also by methods known in the art. The expression of MSF in E. coli using such secretion systems is expected to result in the secretion of the

active polypeptide. This approach has yielded active chimeric antibody fragments [See, e.g., Bitter et al, Science, 240:1041-1043 (1983)].

Alternatively, the MSF may be expressed as a cytoplasmic protein in E. coli, either directly or as a carboxy terminal fusion to proteins, such as thioredoxin, which can maintain many peptides in soluble form in E.coli. The fusion proteins can be cleaved and the free MSF isolated using enzymatic cleavage (enterokinase, Factor Xa) or chemical cleavage (hydroxylamine) depending on the amino acid sequence used to fuse the molecules.

If the cytoplasmic MSF or MSF fusion protein is expressed in inclusion bodies, then either molecule would most likely have to be refolded after complete denaturation with guanidine hydrochloride and a reducing agent a process also known in the art. For procedures for isolation and refolding of intracellularly expressed proteins, see, for example, U. S. Patent No. 4,512,922. If either MSF protein or MSF-fusion protein remain in solution after expression in E. coli, they are likely to not require denaturation but only mild oxidation to generate the correct disulfide bridges.

The compounds expressed through either route in bacterial host cells may then be recovered, purified, and/or characterized with respect to physicochemical, biochemical and/or clinical parameters, all by known methods.

Example 8 - Thioredoxin-MSF Fusion Expression

An MSF can be expressed at high levels in E. coli as a thioredoxin fusion protein as follows. As an example, MSF-K130 was employed. For expression in E. coli, the first 25 amino acids of Exon I, which encode the secretory leader, were removed from the MSF-K130 sequence. An enterokinase site, Asp Asp Asp Asp Lys, was inserted at the 5' end of Exon II of MSF-K130. Additionally, the N-terminal Asp of MSF was deleted and replaced with the dipeptide Asn-Gly, encoded by the sequence AACGGT, which encodes a hydroxylamine cleavage site. The sequence of

MSF-K130 which was added as a fusion to thioredoxin, and which contained certain codon changes for preferred expression in E. coli as shown in Figure 3.

The plasmid expression vector used for expression is illustrated in Figure 4. It contains the following principal features:

Nucleotides 1-2060 contain DNA sequences originating from the plasmid pUC-18 [Norrander et al, Gene 26: 101-106 (1983)] including sequences containing the gene for β-lactamase which confers resistance to the antibiotic ampicillin in host E. coli strains, and a colE1-derived origin of replication. Nucleotides 2061-2221 contain DNA sequences for the major leftward promoter (pL) of bacteriophage λ [Sanger et al, J. Mol. Biol. 162:729-773 (1982)], including three operator sequences, O_L1, O_L2 and O_L3. The operators are the binding sites for λ cI repressor protein, intracellular levels of which control the amount of transcription initiation from pL. Nucleotides 2222-2241 contain a strong ribosome binding sequence derived from that of gene 10 of bacteriophage T7 [Dunn and Studier, J. Mol. Biol. 166:477-535 (1983)].

Nucleotides 2242-2568 contain a DNA sequence encoding the E. coli thioredoxin protein [Lim et al, J. Bacteriol. 163:311-316 (1985)]. There is no translation termination codon at the end of the thioredoxin coding sequence in this plasmid.

Nucleotides 2569-2583 contain DNA sequence encoding the amino acid sequence for a short, hydrophilic, flexible spacer peptide "--GSGSG--". Nucleotides 2584-2598 provide DNA sequence encoding the amino acid sequence for the cleavage recognition site of enterokinase (EC 3.4.4.8), "--DDDDK--" [Maroux et al, J. Biol. Chem. 246:5031-5039 (1971)].

Nucleotides 2599-3132 contain DNA sequence encoding the amino acid sequence of a modified form of mature human IL-11 [Paul et al, Proc. Natl. Acad. Sci. 87:7512-7516 (1990)], deleted for the N-terminal prolyl-residue normally found in the natural protein. The sequence includes a translation

termination codon at the 3'-end of the IL-11 sequence.

Nucleotides 3133-3159 provide a "Linker" DNA sequence containing restriction endonuclease sites. Nucleotides 3160-3232 provide a transcription termination sequence based on that of the E. coli aspA gene [Takagi et al, Nucl. Acids Res. 13:2063-2074 (1985)]. Nucleotides 3233-3632 are DNA sequences derived from pUC-18.

This plasmid is modified in the following manner to replace the ribosome binding site of bacteriophage T7 with that of λ CII. In the above-described plasmid, nucleotides 2222 and 2241 were removed by conventional means. Inserted in place of those nucleotides was a sequence of nucleotides formed by nucleotides 35566 to 35472 and 38137 to 38361 from bacteriophage lambda as disclosed and described in Sanger et al (1982) cited above.

The DNA sequence encoding human IL11 in modified pALtrxA/EK/IL11Δ Pro-581 (nucleotides 2599-3135) is replaced by the sequence shown in Figure 3.

The resulting plasmid was transformed into the E. coli host strain GI724 (F^-, $lacI^q$, $lacP^{L8}$, ampC::λ ci^+) by the procedure of Dagert and Ehrlich, Gene 6: 23 (1979). The untransformed host strain E. coli GI724 was deposited with the American Type Culture Collection, 12301 Parklawn Drive, Rockville, Maryland on January 31, 1991 under ATCC No. 55151 for patent purposes pursuant to applicable laws and regulations. Transformants were selected on 1.5% w/v agar plates containing IMC medium, which is composed of M9 medium [Miller, "Experiments in Molecular Genetics", Cold Spring Harbor Laboratory, New York (1972)] supplemented with 0.5% w/v glucose, 0.2% w/v casamino acids and 100 μg/ml ampicillin.

GI724 contains a copy of the wild-type λ cI repressor gene stably integrated into the chromosome at the ampC locus, where it has been placed under the transcriptional control of Salmonella typhimurium trp promoter/operator sequences. In GI724, λ cI protein is made only during growth in tryptophan-free media, such as minimal media or a minimal medium supplemented

with casamino acids such as IMC, described above. Addition of tryptophan to a culture of GI724 will repress the trp promoter and turn off synthesis of λcI, gradually causing the induction of transcription from pL promoters if they are present in the cell.

GI724 transformed with the MSF containing plasmid was grown at 37°C to an A_{550} of 0.5 in IMC medium. Tryptophan was added to a final concentration of 100 μg/ml and the culture incubated for a further 4 hours.

All of the thioredoxin-MSF fusion protein was found in the soluble cellular fraction, representing up to 10% of the total protein. The fusion protein was heat stable, remaining soluble after treatment at 80 degrees Celsius for fifteen minutes. The fusion protein has shown biological activity in the fibrin clot assay described in Example 10.

Example 9 - Other Expression Systems

Manipulations can be performed for the construction of an insect vector for expression of MSF polypeptides in insect cells [See, e.g., procedures described in published European patent application 155,476].

Similarly yeast vectors may be constructed employing yeast regulatory sequences to express cDNA encoding the precursor, in yeast cells to yield secreted extracellular active MSF. Alternatively the polypeptide may be expressed intracellularly in yeast, the polypeptide isolated and refolded to yield active MSF. [See, e.g., procedures described in PCT Publication WO 86/00639 and European Patent Publication EP 123,289.]

Example 10 - Biological Activities of Human MSFs

The following assays were performed using the purified native urinary Meg-CSF described in Example 1, and crude or highly purified preparations of recombinant MSF-K130. The other recombinant or naturally occurring MSFs may be tested for MSF biological properties and activity in these same assays

following the teachings herein. Alternatively, other assays known in the art may be used to test the MSFs of this invention for biological activity.

A. Murine Fibrin Clot Assay

The Meg-CSF obtained from Step 7 of the purification of Example 1 was tested for activity in the megakaryocyte colony formation assay performed substantially as described in S. Kuriya et al, Exp. Hematol. 15:896-901 (1987). A fibrin clot was formed containing 2.5×10^5 murine bone marrow cells in a 6-well plate. The diluted sample was layered around the clot and incubated for 6 days. Thereafter, cells were fixed and megakaryocytes were stained for acetylcholinesterase, a specific marker for murine megakaryocytes. A colony was defined as three or more megakaryocytes per unit area.

A mixture of pure and mixed colonies containing megakaryocyte colonies were routinely observed: 70% pure megakaryocyte colonies containing no additional cell types: 30 % mixed megakaryocyte colonies containing additional non-megakaryocyte cell types to 50% pure: 50% mixed type depending on the assay. The pure colonies typically contained on average 4 to 5 cells per colony, ranging from 3 to 8 cells per colony. The cells within the colony are variable in size and appear to contain both mature and immature megakaryocytes. The megakaryocytes were fairly dispersed within the colony. A typical mixed megakaryocyte colony is composed on average of 10 cells per colony ranging from 7 to 17 cells. The cells appear more clustered than the megakaryocytes in pure megakaryocyte colonies.

The following control samples were included in every assay. A positive control was WEHI conditioned medium (murine IL-3), which produced between 7-25 (average 12) megakaryocyte colonies per clot, approximately 50% pure and 50% mixed megakaryocyte colonies. Another positive control was serum taken from lethally irradiated dogs at the nadir or low point of the platelet count [see Mazur et al, Exp. Hematol. 13:1164-1172

(1985)], which produced between 6-22 (average 15) megakaryocyte colonies per clot, of which approximately 70% were pure and 30% were mixed megakaryocyte colonies. The negative control was Iscoves Medium, which produced 2-4 megakaryocyte colonies per clot. In the assay, the purified urinary Meg-CSF has a specific activity of greater than approximately $5X10^7$ dilution units/mg of protein. A unit of activity is defined as described in Example 1.

The major Meg-CSF obtained from bone marrow transplant urine eluted from the S-Toyopearl cation exchange column chromatography step in the purification of Example 1 has been analyzed in this assay alone, together, and in combination with other cytokines. In the fibrin clot assay, it produced between 6-16 (average 13) megakaryocyte colonies, with 50-70% pure megakaryocyte colonies. The urinary Meg-CSF has been shown to have variable synergy with murine IL-3 and does not synergize with human IL-6 or murine IL-4 in the fibrin clot culture system.

Megakaryocyte colony formation was observed in response to recombinant MSF-K130 and in response to recombinant 4aA (MSF-L924) in the murine bone marrow fibrin clot assay. Murine megakaryocytes were identified as acetylcholinesterase positive cells and a megakaryocyte colony was defined as greater than three megakaryocyte cells per unit area in a fibrin clot culture. Recombinant MSF typically stimulated megakaryocyte colonies of three to six cells/unit area and averaged between 6 to 15 colonies/2.5 X 10^5 murine bone marrow cells.

Two types of megakaryocyte colonies were observed in the assay, pure megakaryocyte colonies and megakaryocyte cells with other cell types, termed mixed megakaryocyte colonies. In one fibrin clot, the two colony types were at a ratio between 1:1 to 7:3 pure colonies to mixed megakaryocyte colonies. This ratio was consistent throughout the purification of recombinant MSF-K130. The number of megakaryocyte cells/colony and size of megakaryocytes were about the same for both pure and mixed

colonies,. some megakaryocytes were smaller in the mixed megakaryocyte colonies.

An increase in bioactivity was usually observed from active MSF fractions obtained from the C4 RP-HPLC column. All three pools from the S Toyopearl cation exchange column yielded bioactive MSF protein on RP-HPLC. The final specific activity of MSF after the RP-HPLC step was greater than 1×10^7 dilution units/mg in all three pools. The active peaks were also positive on the MSF Western Blot.

When RP-HPLC-purified MSF-K130 from the A pool was subjected to SDS-PAGE under non-reducing conditions, bioactive protein was extracted from gel slices corresponding to 35-50 kD molecular weight species. A silver stain gel and Western immunoblot data showed that 95% of the 35-50 kD recombinant MSF protein reduced to 18-21 kD and 5% did not shift upon reduction on a 10-20% acrylamide gradient SDS-PAGE.

The supernatant from COS-1 cells transfected with MSF-K130 cDNA was variably active on the fibrin clot assay. In each assay the samples were tested in duplicate and in three dilutions.

B. Human Plasma Clot Megakaryocyte Colony Formation

The human urinary Meg-CSF of this invention was also tested for human activity on the plasma clot MSF assay described in E. Mazur et al, Blood 57:277-286 (1981) with modifications. Non-adherent peripheral blood cells were isolated from Leukopacs and frozen in aliquots. The test sample was mixed with platelet-poor human AB plasma and 1.25×10^5 cells in 24-well plates and allowed to clot by the addition of calcium. After a 12 day incubation, megakaryocytes were identified using a monoclonal antibody directed against platelet glycoproteins IIb/IIIa and a horseradish peroxidase/anti-peroxidase chromogenic detection system. Recombinant human IL-3 [Genetics Institute, Inc.] was used as a positive control, producing 12-30 megakaryocyte colonies per clot with approximately 60% pure and 40% mixed megakaryocyte colonies. As in the murine assay, the aplastic

dog serum was also used as a positive control, which produced between 5-10 megakaryocyte colonies per clot, of which approximately 50% were pure megakaryocyte colonies containing less than 10 cells, and 50% were mixed megakaryocyte colonies containing more than 40 megakaryocytes. The negative control was Alpha Medium, which produced 0-1 megakaryocyte colonies per clot.

The human urinary Meg-CSF product from Step 6 of the purification scheme of Example 1 had variable activity in this assay. MSF-K130 has shown variable activity in the human plasma clot megakaryocyte colony assay.

C. Synergistic Effects

Recombinant MSF-K130 COS-1 cell supernatant and purified recombinant MSF were assayed alone and in combination with other cytokines in the various CFU-MEG assay systems, fibrin clot, agar and the human CFU-MEG plasma clot assays.

Variable synergy with IL-3 was observed in the murine bone marrow fibrin clot assay. The stimulation of megakaryocyte colonies increased above either protein alone when both murine IL-3 and MSF-K130 or MSF 4aA (MSF-L924) were cultured with bone marrow cells progenitors in the fibrin clot assay. A suboptimal level of murine IL-3 (15 units/ml) and an optimal level of MSF-K130 each stimulate an average of 6-15 CFU-meg/2.5 X 10^5 murine bone marrow cells in the fibrin clot assay. In combination, increased megakaryocyte colony stimulation of over 35 megakaryocyte colonies have been observed. The ratio of pure megakaryocyte colonies to mixed megakaryocyte colonies and the size of the megakaryocyte colonies were the same for the combination cultures as for the individual MSF cultures.

D. E. Coli Expressed MSF Activity

MSF expressed in Escherichia coli as a thioredoxin-MSF-K130 fusion protein was soluble and active in the fibrin clot assay. E. coli expressed MSF-K130 stimulated the same range of CFU-meg/2.5X10^5 murine bone marrow cells as COS derived MSF-K130. This activity was not neutralized by the addition of anti-IL-3

antibody at a level that did neutralize CFU-Meg formation by IL-3. Megakaryocyte colony formation activity of the MSF-K130 thioredoxin fusion protein from E. coli lysate was $5X10^6$ dilution units/ml. The specific activity of the MSF-K130 thioredoxin fusion protein in E. coli lysate was greater than $1X10^6$ U/mg. Thioredoxin was not active in the assay.

Example 11 - Construction of CHO Cell Lines Expressing High Levels of MSF

One method for producing high levels of the MSF protein of the invention from mammalian cells involves the construction of cells containing multiple copies of the cDNA encoding the MSF.

The cDNA is co-transfected with an amplifiable marker, e.g., the DHFR gene for which cells containing increasing concentrations of methotrexate (MTX) according to the procedures of Kaufman and Sharp, J. Mol. Biol. (1982) supra. This approach can be employed with a number of different cell types. Alternatively, the MSF cDNA and drug resistance selection gene (e.g., DHFR) may be introduced into the same vector. One desirable vector for this approach is pED4DPC-1. MSF-K130 and MSF-N141 are being expressed in vector pEMC3-1, a vector identical to pE04DPC-1, but in which the polylinker has been changed (PstI, NotI, SalI, SnaBI, EcoRI, PacI) as described above per pMT21.

For example, the pMT21 vector containing the MSF gene in operative association with other plasmid sequences enabling expression thereof is introduced into DHFR-deficient CHO cells, DUKX-BII, along with a DHFR expression plasmid such as pAdD26SVpA3 [Kaufman, Proc. Natl. Acad. Sci. 82:689-693 (1985)] by calcium phosphate co-precipitation and transfection.

Alternatively, the pED4DPC-1 vector containing the MSF gene in operative association with other plasmid sequences enabling expression thereof is introduced into DHFR-deficient CHO cells, DUKX-BII, by protoplast fusion or transfection. The MSF gene and DHFR marker gene are both efficiently expressed when MSF is

introduced into pEMC2B1. DHFR expressing transformants are selected for growth in alpha media with dialyzed fetal calf serum. Transformants are checked for expression of MSF by Western blotting, bioassay, or RNA blotting and positive pools are subsequently selected for amplification by growth in increasing concentrations of MTX (sequential steps in 0.02, 0.2, 1.0 and 5uM MTX) as described in Kaufman et al., Mol. Cell Biol., 5:1750 (1983). The amplified lines are cloned, and MSF protein expression is monitored by the fibrin clot assay. MSF expression is expected to increase with increasing levels of MTX resistance.

In any of the expression systems described above, the resulting cell lines can be further amplified by appropriate drug selection, resulting cell lines recloned and the level of expression assessed using the murine fibrin clot assay described in Example 10.

The MSF expressing CHO cell lines can be adapted for growth in serum-free medium. MSF expressed in CHO cells is purified from serum-free conditioned medium using the same purification scheme as COS-1 cell supernatant. Homogeneous MSF can be isolated from conditioned medium from the cell line using methods familiar in the art, including techniques such as lectin-affinity chromatography, reverse phase HPLC, FPLC and the like.

The foregoing descriptions detail presently preferred embodiments of the invention. Numerous modifications and variations in practice of this invention are expected to occur to those skilled in the art. Such modifications and variations are encompassed within the following claims.

WHAT IS CLAIMED IS:

1. An MSF protein, substantially free from association with other proteinaceous materials and contaminants with which it is associated in natural sources, said protein comprising the amino acid sequence of Exon II, Exon III and Exon IV, of Figure 1 and having an amino terminal sequence encoding a secretory leader and initiating methionine preceding Exon II and a termination codon following Exon IV said protein being characterized by the ability to stimulate growth and development of colonies of megakaryocyte cells.

2. A protein according to claim 1 wherein said amino terminal sequence is Exon I of Figure 1.

3. A protein according to claim 2, additionally comprising the amino acid sequences of Exon XII and at least one of Exons V, VI, VII, VIII, IX, X and XI of Figure 1, wherein said termination codon is present in Exon XII.

4. An MSF protein, substantially free from association with other proteinaceous materials and contaminants with which it is associated in natural sources, comprising an amino acid sequence selected from the group consisting of

(a) a contiguous amino acid sequence comprising amino acids 1-25 of Figure 1 fused in frame to amino acids 67-106 of Figure 1, fused in frame to amino acids 200-250 of Figure 1;

(b) a contiguous amino acid sequence comprising amino acids 1-106 of Figure 1 fused in frame to amino acids 200-250 of Figure 1;

(c) a contiguous amino acid sequence comprising amino acids 1-156 of Figure 1 fused in frame to amino acids 200-250 of Figure 1;

(d) a contiguous amino acid sequence comprising amino acids 1-25 fused in frame to amino acids 67-106, fused in frame

to amino acids 200-250;

(e) a contiguous amino acid sequence comprising amino acids 1-25 of Figure 1 fused in frame to amino acids 67-156 of Figure 1; and

(f) the sequence from amino acid 1 through amino acid 130 of Figure 1;

(g) the sequence from amino acid 1 through amino acid 141 of Figure 1;

(h) the sequence from amino acid 1 through 156 of Figure 1;

(i) the sequence from amino acid 1 through amino acid 172 of Figure 1;

(j) the sequence from amino acid 1 through amino acid 192 of Figure 1;

(k) the sequence from amino acid 1 through amino acid 204 of Figure 1;

(l) the sequence from amino acid 1 through amino acid 209 of Figure 1;

(m) the sequence from amino acid 1 through amino acid 220 of Figure 1;

(n) homodimers or heterodimers of sequences (a) through (m).

5. An MSF DNA sequence selected from the group consisting of

(a) a DNA sequence comprising nucleotides 1 through 390 of Figure 1;

(b) a DNA sequence comprising nucleotides 1 through 423 of Figure 1;

(c) a DNA sequence comprising nucleotides 1 through 516 of Figure 1;

(d) a DNA sequence comprising nucleotides 1 through 576 of Figure 1;

(e) a DNA sequence comprising nucleotides 1 through 612 of Figure 1;

(f) a DNA sequence comprising nucleotides 1 through 627 of

Figure 1;

(g) a DNA sequence comprising nucleotides 1 through 660 of Figure 1;

(h) DNA sequences encoding homodimers or heterodimers of sequences (a) through (g);

(i) allelic variations of the sequences of (a) through (g); and

(j) a DNA sequence capable of hybridizing to any of (a) through (i), which encodes a peptide or polypeptide having activity in the fibrin clot assay.

6. An MSF DNA sequence comprising a 5' sequence selected from the group consisting of

(a) a DNA sequence comprising nucleotides 1-76 of Figure 1 fused in frame to nucleotides 200-319 of Figure 1, fused in frame to nucleotides 598-748 of Figure 1;

(b) a DNA sequence comprising nucleotides 1-319 of Figure 1 fused in frame to nucleotides 598-748 of Figure 1;

(c) a DNA sequence comprising nucleotides 1-469 of Figure 1 fused in frame to nucleotides 598-748 of Figure 1;

(d) a DNA sequence comprising nucleotides 1-76 of Figure 1 fused in frame to nucleotides 200-319 of Figure 1, fused in frame to nucleotides 598-748 of Figure 1;

(e) the sequence from nucleotides 1 through 469 of Figure 1;

(f) a nucleotide sequence comprising nucleotides 1 to 76 of Figure 1 fused in frame to nucleotides 200 through 469 of Figure 1;

(g) DNA sequences encoding homodimers or heterodimers of sequences (a) through (f);

(h) allelic variations of the sequences of (a) through (f); and

(i) a DNA sequence capable of hybridizing to any of (a) through (h), which encodes a peptide or polypeptide having activity in the fibrin clot assay.

7. A process for producing an MSF protein comprising

(a) culturing in a culture medium a cell line transformed with a DNA sequence of claim 5 or 6 encoding expression of an MSF protein in operative association with an expression control sequence therefor; and

(b) recovering said MSF protein from said culture medium.

8. An MSF protein produced by the process of claim 7.

9. A cell transformed with an MSF DNA sequence of claim 5 or 6 in operative association with an expression control sequence.

10. A pharmaceutical composition comprising a therapeutically effective amount of an MSF protein of claims 1, 2, 3 or 4 thereof in a pharmaceutically effective vehicle.

FIGURE 1

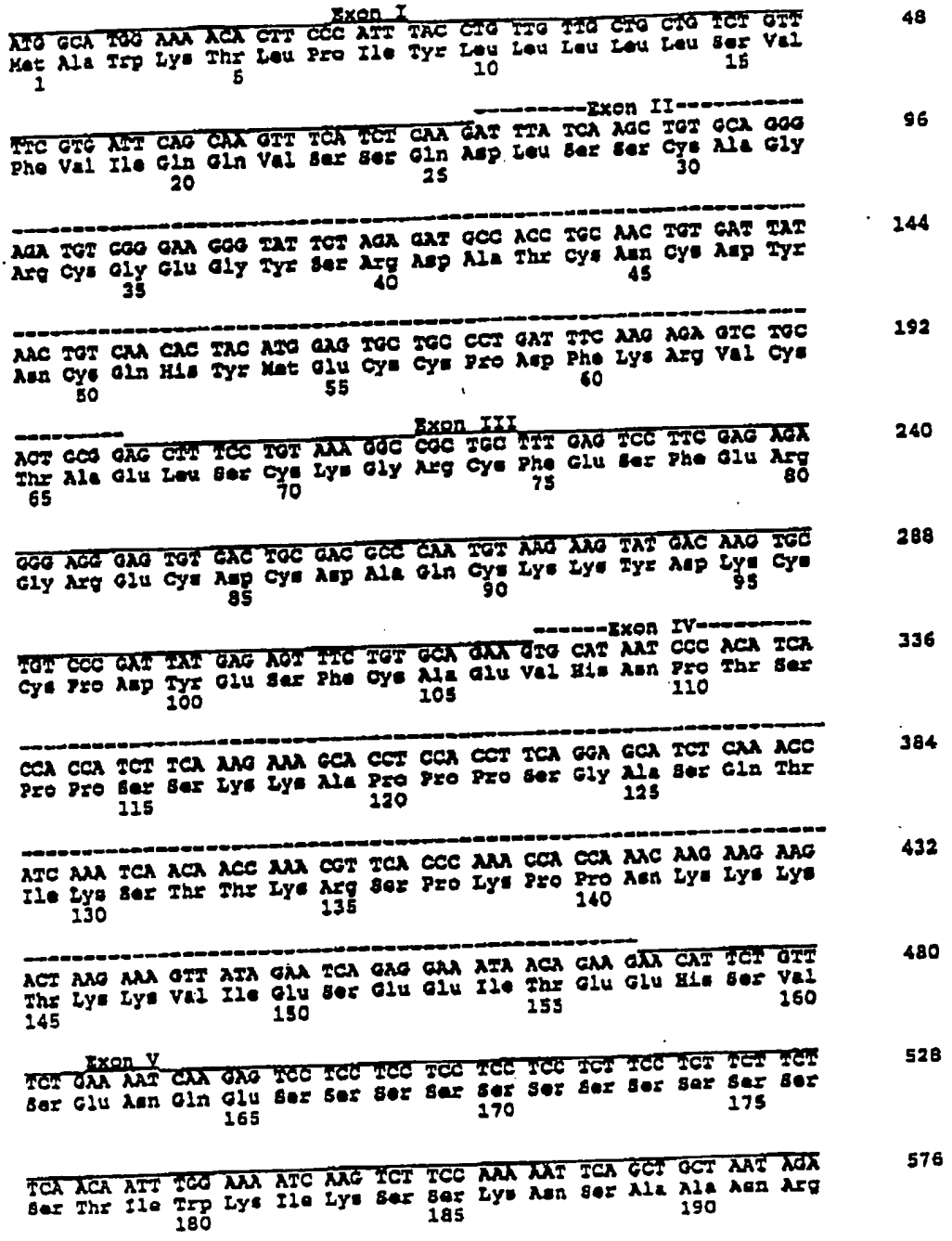
Exon I
ATG GCA TGG AAA ACA CTT CCC ATT TAC CTG TTG TTG CTG CTG TCT GTT 48
Met Ala Trp Lys Thr Leu Pro Ile Tyr Leu Leu Leu Leu Leu Ser Val
1 5 10 15
Exon II
TTC GTC ATT CAG CAA GTT TCA TCT CAA GAT TTA TCA AGC TGT GCA GGG 96
Phe Val Ile Gln Gln Val Ser Ser Gln Asp Leu Ser Ser Cys Ala Gly
20 25 30
AGA TGT CGG GAA GGG TAT TCT AGA GAT GCC ACC TGC AAC TGT GAT TAT 144
Arg Cys Gly Glu Gly Tyr Ser Arg Asp Ala Thr Cys Asn Cys Asp Tyr
35 40 45
AAC TGT CAA CAC TAC ATG GAG TGC TGC CCT GAT TTC AAG AGA GTC TGC 192
Asn Cys Gln His Tyr Met Glu Cys Cys Pro Asp Phe Lys Arg Val Cys
50 55 60
Exon III
ACT GCG GAG CTT TCC TGT AAA GGC CGC TGC TTT GAG TCC TTC GAG AGA 240
Thr Ala Glu Leu Ser Cys Lys Gly Arg Cys Phe Glu Ser Phe Glu Arg
65 70 75 80
GGG AGG GAG TGT GAC TGC GAC GCC CAA TGT AAG AAG TAT GAC AAG TGC 288
Gly Arg Glu Cys Asp Cys Asp Ala Gln Cys Lys Lys Tyr Asp Lys Cys
85 90 95
Exon IV
TGT CCC GAT TAT GAG AGT TTC TGT GCA GAA GTG CAT AAT CCC ACA TCA 336
Cys Pro Asp Tyr Glu Ser Phe Cys Ala Glu Val His Asn Pro Thr Ser
100 105 110
CCA CCA TCT TCA AAG AAA GCA CCT CCA CCT TCA GGA GCA TCT CAA ACC 384
Pro Pro Ser Ser Lys Lys Ala Pro Pro Pro Ser Gly Ala Ser Gln Thr
115 120 125
ATC AAA TCA ACA ACC AAA CGT TCA CCC AAA CCA CCA AAC AAG AAG AAG 432
Ile Lys Ser Thr Thr Lys Arg Ser Pro Lys Pro Pro Asn Lys Lys Lys
130 135 140
ACT AAG AAA GTT ATA GAA TCA GAG GAA ATA ACA GAA GAA CAT TCT GTT 480
Thr Lys Lys Val Ile Glu Ser Glu Glu Ile Thr Glu Glu His Ser Val
145 150 155 160
Exon V
TCT GAA AAT CAA GAG TCC TCC TCC TCC TCC TCC TCT TCC TCT TCT TCT 528
Ser Glu Asn Gln Glu Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
165 170 175
TCA ACA ATT TGG AAA ATC AAG TCT TCC AAA AAT TCA GCT GCT AAT AGA 576
Ser Thr Ile Trp Lys Ile Lys Ser Ser Lys Asn Ser Ala Ala Asn Arg
180 185 190

FIGURE 1 (Cont'd.)

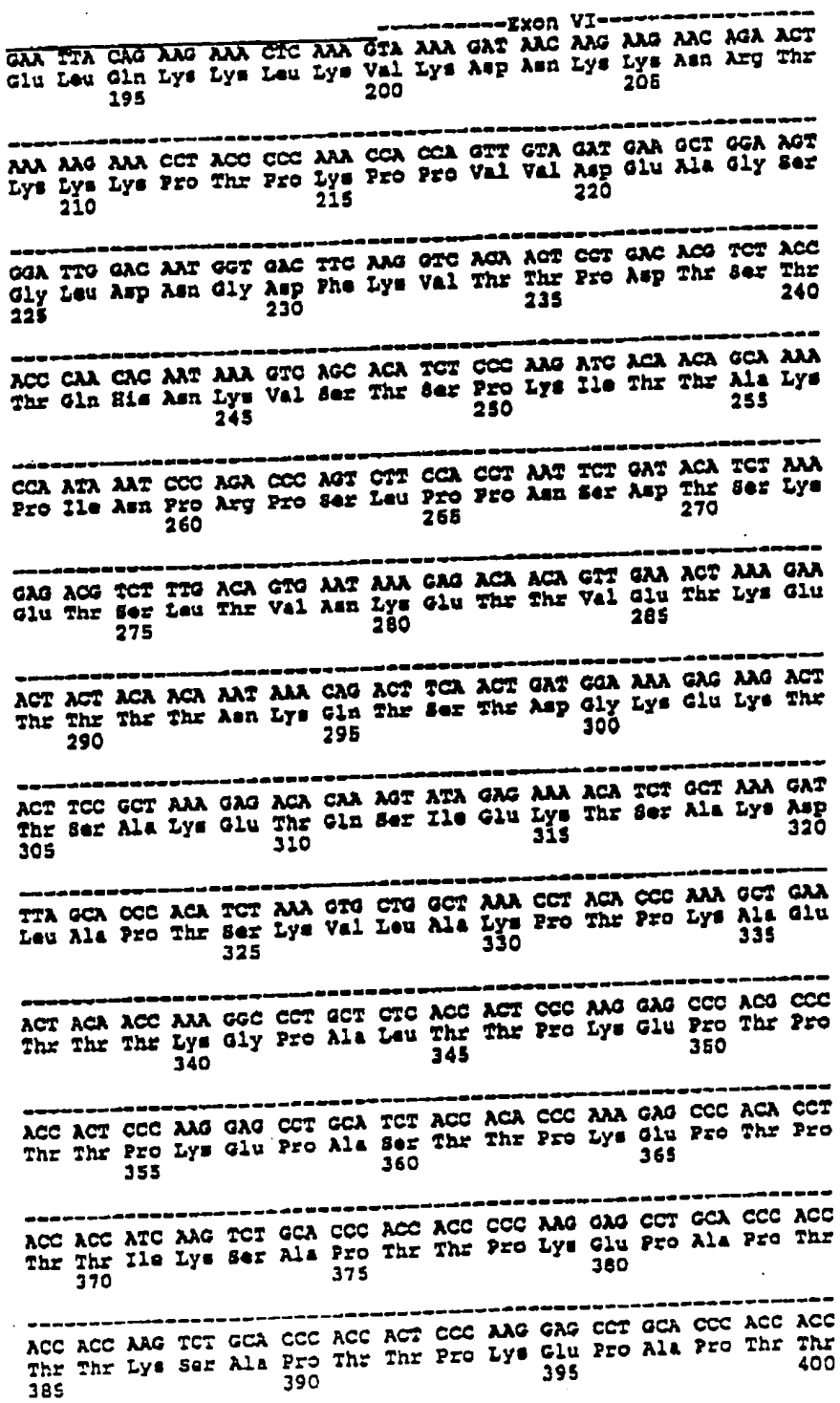

FIGURE 1 (Cont'd.)

```
ACC AAG GAG CCT GCA CCC ACC ACT CCC AAG GAG CCT GCA CCC ACC ACC                1248
Thr Lys Glu Pro Ala Pro Thr Thr Pro Lys Glu Pro Ala Pro Thr Thr
                405                 410                 415

ACC AAG GAG CCT GCA CCC ACC ACC ACC AAG TCT GCA CCC ACC ACT CCC               1296
Thr Lys Glu Pro Ala Pro Thr Thr Thr Lys Ser Ala Pro Thr Thr Pro
            420                 425                 430

AAG GAG CCT GCA CCC ACC ACC CCC AAG AAG CCT GCC CCA ACT ACC CCC               1344
Lys Glu Pro Ala Pro Thr Thr Pro Lys Lys Pro Ala Pro Thr Thr Pro
        435                 440                 445

AAG GAG CCT GCA CCC ACC ACT CCC AAG GAG CCT ACA CCC ACC ACT CCC               1392
Lys Glu Pro Ala Pro Thr Thr Pro Lys Glu Pro Thr Pro Thr Thr Pro
    450                 455                 460

AAG GAG CCT GCA CCC ACC ACC AAG GAG CCT GCA CCC ACC ACT CCC AAA               1440
Lys Glu Pro Ala Pro Thr Thr Lys Glu Pro Ala Pro Thr Thr Pro Lys
465                 470                 475                 480

GAG CCT GCA CCC ACT GCC CCC AAG AAG CCT GCC CCA ACT ACC CCC AAG               1488
Glu Pro Ala Pro Thr Ala Pro Lys Lys Pro Ala Pro Thr Thr Pro Lys
                485                 490                 495

GAG CCT GCA CCC ACC ACT CCC AAG GAG CCT GCA CCC ACC ACC ACC AAG               1536
Glu Pro Ala Pro Thr Thr Pro Lys Glu Pro Ala Pro Thr Thr Thr Lys
            500                 505                 510

GAG CCT TCA CCC ACC ACT CCC AAG GAG CCT GCA CCC ACC ACC ACC AAG               1584
Glu Pro Ser Pro Thr Thr Pro Lys Glu Pro Ala Pro Thr Thr Thr Lys
        515                 520                 525

TCT GCA CCC ACC ACT ACC AAG GAG CCT GCA CCC ACC ACT ACC AAG TCT               1632
Ser Ala Pro Thr Thr Thr Lys Glu Pro Ala Pro Thr Thr Thr Lys Ser
    530                 535                 540

GCA CCC ACC ACT CCC AAG GAG CCT TCA CCC ACC ACC ACC AAG GAG CCT               1680
Ala Pro Thr Thr Pro Lys Glu Pro Ser Pro Thr Thr Thr Lys Glu Pro
545                 550                 555                 560

GCA CCC ACC ACT CCC AAG GAG CCT GCA CCC ACC ACC CCC AAG AAG CCT               1728
Ala Pro Thr Thr Pro Lys Glu Pro Ala Pro Thr Thr Pro Lys Lys Pro
                565                 570                 575

GCC CCA ACT ACC CCC AAG GAG CCT GCA CCC ACC ACT CCC AAG GAA CCT               1776
Ala Pro Thr Thr Pro Lys Glu Pro Ala Pro Thr Thr Pro Lys Glu Pro
            580                 585                 590

GCA CCC ACC ACC ACC AAG AAG CCT GCA CCC ACC GCT CCC AAA GAG CCT               1824
Ala Pro Thr Thr Thr Lys Lys Pro Ala Pro Thr Ala Pro Lys Glu Pro
        595                 600                 605
```

FIGURE 1 (Cont'd.)

```
GCC CCA ACT ACC CCC AAG GAG ACT GCA CCC ACC ACC CCC AAG AAG CTC      1872
Ala Pro Thr Thr Pro Lys Glu Thr Ala Pro Thr Thr Pro Lys Lys Leu
    610                 615                 620

ACG CCC ACC ACC CCC GAG AAG CTC GCA CCC ACC ACC CCT GAG AAG CCC      1920
Thr Pro Thr Thr Pro Glu Lys Leu Ala Pro Thr Thr Pro Glu Lys Pro
625                 630                 635                 640

GCA CCC ACC ACC CCT GAG GAG CTC GCA CCC ACC ACC CCT GAG GAG CCC      1968
Ala Pro Thr Thr Pro Glu Glu Leu Ala Pro Thr Thr Pro Glu Glu Pro
                645                 650                 655

ACA CCC ACC ACC CCT GAG GAG CCT GCT CCC ACC ACT CCC AAG GCA GCG      2016
Thr Pro Thr Thr Pro Glu Glu Pro Ala Pro Thr Thr Pro Lys Ala Ala
            660                 665                 670

GCT CCC AAC ACC CCT AAG GAG CCT GCT CCA ACT ACC CCT AAG GAG CCT      2064
Ala Pro Asn Thr Pro Lys Glu Pro Ala Pro Thr Thr Pro Lys Glu Pro
        675                 680                 685

GCT CCA ACT ACC CCT AAG GAG CCT GCT CCA ACT ACC CCT AAG GAG ACT      2112
Ala Pro Thr Thr Pro Lys Glu Pro Ala Pro Thr Thr Pro Lys Glu Thr
    690                 695                 700

GCT CCA ACT ACC CCT AAA GGG ACT GCT CCA ACT ACC CTC AAG GAA CCT      2160
Ala Pro Thr Thr Pro Lys Gly Thr Ala Pro Thr Thr Leu Lys Glu Pro
705                 710                 715                 720

GCA CCC ACT ACT CCC AAG AAG CCT GCC CCC AAG GAG CTT GCA CCC ACC      2208
Ala Pro Thr Thr Pro Lys Lys Pro Ala Pro Lys Glu Leu Ala Pro Thr
                725                 730                 735

ACC ACC AAG GAG CCC ACA TCC ACC ACC TCT GAC AAG CCC GCT CCA ACT      2256
Thr Thr Lys Glu Pro Thr Ser Thr Thr Ser Asp Lys Pro Ala Pro Thr
            740                 745                 750

ACC CCT AAG GGG ACT GCT CCA ACT ACC CCT AAG GAG CCT GCT CCA ACT      2304
Thr Pro Lys Gly Thr Ala Pro Thr Thr Pro Lys Glu Pro Ala Pro Thr
        755                 760                 765

ACC CCT AAG GAG CCT GCT CCA ACT ACC CCT AAG GGG ACT GCT CCA ACT      2352
Thr Pro Lys Glu Pro Ala Pro Thr Thr Pro Lys Gly Thr Ala Pro Thr
    770                 775                 780

ACC CTC AAG GAA CCT GCA CCC ACT ACT CCC AAG AAG CCT GCC CCC AAG      2400
Thr Leu Lys Glu Pro Ala Pro Thr Thr Pro Lys Lys Pro Ala Pro Lys
785                 790                 795                 800

GAG CTT GCA CCC ACC ACC ACC AAG GGG CCC ACA TCC ACC ACC TCT GAC      2448
Glu Leu Ala Pro Thr Thr Thr Lys Gly Pro Thr Ser Thr Thr Ser Asp
                805                 810                 815
```

FIGURE 1 (Cont'd.)

```
                                                                    2496
AAG CCT GCT CCA ACT ACA CCT AAG GAG ACT GCT CCA ACT ACC CCC AAG
Lys Pro Ala Pro Thr Thr Pro Lys Glu Thr Ala Pro Thr Thr Pro Lys
        820                     825                 830

                                                                    2544
GAG CCT GCA CCC ACT ACC CCC AAG AAG CCT GCT CCA ACT ACT CCT GAG
Glu Pro Ala Pro Thr Thr Pro Lys Lys Pro Ala Pro Thr Thr Pro Glu
    835                     840                 845

                                                                    2592
ACA CCT CCT CCA ACC ACT TCA GAG GTC TCT ACT CCA ACT ACC ACC AAG
Thr Pro Pro Pro Thr Thr Ser Glu Val Ser Thr Pro Thr Thr Thr Lys
850                     855                 860

                                                                    2640
GAG CCT ACC ACT ATC CAC AAA AGC CCT GAT GAA TCA ACT CCT GAG CTT
Glu Pro Thr Thr Ile His Lys Ser Pro Asp Glu Ser Thr Pro Glu Leu
865                     870                 875                 880

                                                                    2688
TCT GCA GAA CCC ACA CCA AAA GCT CTT GAA AAC AGT CCC AAG GAA CCT
Ser Ala Glu Pro Thr Pro Lys Ala Leu Glu Asn Ser Pro Lys Glu Pro
                885                     890                 895

                                                                    2736
GGT GTA CCT ACA ACT AAG ACT CCT GCA GCG ACT AAA CCT GAA ATG ACT
Gly Val Pro Thr Thr Lys Thr Pro Ala Ala Thr Lys Pro Glu Met Thr
        900                     905                 910

                                                                    2784
ACA ACA GCT AAA GAC AAG ACA ACA GAA AGA GAC TTA CGT ACT ACA CCT
Thr Thr Ala Lys Asp Lys Thr Thr Glu Arg Asp Leu Arg Thr Thr Pro
    915                     920                 925

                                                                    2832
GAA ACT ACA ACT GCT GCA CCT AAG ATG ACA AAA GAG ACA GCA ACT ACA
Glu Thr Thr Thr Ala Ala Pro Lys Met Thr Lys Glu Thr Ala Thr Thr
930                     935                 940

                                                                    2880
ACA GAA AAA ACT ACC GAA TCC AAA ATA ACA GCT ACA ACC ACA CAA GTA
Thr Glu Lys Thr Thr Glu Ser Lys Ile Thr Ala Thr Thr Thr Gln Val
945                     950                 955                 960

                                                                    2928
ACA TCT ACC ACA ACT CAA GAT ACC ACA CCA TTC AAA ATT ACT ACT CTT
Thr Ser Thr Thr Thr Gln Asp Thr Thr Pro Phe Lys Ile Thr Thr Leu
                965                     970                 975

                                                                    2976
AAA ACA ACT ACT CTT GCA CCC AAA GTA ACT ACA ACA AAA AAG ACA ATT
Lys Thr Thr Thr Leu Ala Pro Lys Val Thr Thr Thr Lys Lys Thr Ile
        980                     985                 990

                                                                    3024
ACT ACC ACT GAG ATT ATG AAC AAA CCT GAA GAA ACA GCT AAA CCA AAA
Thr Thr Thr Glu Ile Met Asn Lys Pro Glu Glu Thr Ala Lys Pro Lys
    995                     1000                1005

                                                                    3072
GAC AGA GCT ACT AAT TCT AAA GCG ACA ACT CCT AAA CCT CAA AAG CCA
Asp Arg Ala Thr Asn Ser Lys Ala Thr Thr Pro Lys Pro Gln Lys Pro
1010                    1015                1020
```

FIGURE 1 (Cont'd.)

```
                                                                        3120
ACC AAA CCA CCC AAA AAA CCC ACT TCT ACC AAA AAG CCA AAA ACA ATG
Thr Lys Ala Pro Lys Lys Pro Thr Ser Thr Lys Lys Pro Lys Thr Met
1025                1030                1035                1040

                                                                        3168
CCT AGA GTG AGA AAA CCA AAG ACG ACA CCA ACT CCC CGC AAG ATG ACA
Pro Arg Val Arg Lys Pro Lys Thr Thr Pro Thr Pro Arg Lys Met Thr
                    1045                1050                1055

                                                                        3216
TCA ACA ATG CCA GAA TTG AAC CCT ACC TCA AGA ATA GCA GAA GCC ATG
Ser Thr Met Pro Glu Leu Asn Pro Thr Ser Arg Ile Ala Glu Ala Met
            1060                1065                1070

                                                                        3264
CTC CAA ACC ACC ACC AGA CCT AAC CAA ACT CCA AAC TCC AAA CTA GTT
Leu Gln Thr Thr Thr Arg Pro Asn Gln Thr Pro Asn Ser Lys Leu Val
        1075                1080                1085

                                                                        3312
GAA GTA AAT CCA AAG AGT GAA GAT GCA GGT GGT GCT GAA GGA GAA ACA
Glu Val Asn Pro Lys Ser Glu Asp Ala Gly Gly Ala Glu Gly Glu Thr
    1090                1095                1100

                                                                        3360
CCT CAT ATG CTT CTC AGG CCC CAT GTG TTC ATG CCT GAA GTT ACT CCC
Pro His Met Leu Leu Arg Pro His Val Phe Met Pro Glu Val Thr Pro
1105                1110                1115                1120

                                                                        3408
GAC ATG GAT TAC TTA CCG AGA GTA CCC AAT CAA GGC ATT ATC ATC AAT
Asp Met Asp Tyr Leu Pro Arg Val Pro Asn Gln Gly Ile Ile Ile Asn
                1125                1130                1135

                                    Exon VII
                                                                        3456
CCC ATG CTT TCC GAT GAG ACC AAT ATA TGC AAT GGT AAG CCA GTA GAT
Pro Met Leu Ser Asp Glu Thr Asn Ile Cys Asn Gly Lys Pro Val Asp
            1140                1145                1150

                                                                        3504
GGA CTG ACT ACT TTG CGC AAT GGG ACA TTA GTT GCA TTC CGA GGT CAT
Gly Leu Thr Thr Leu Arg Asn Gly Thr Leu Val Ala Phe Arg Gly His
        1155                1160                1165

----------Exon VIII----------                                          3552
TAT TTC TGG ATG CTA AGT CCA TTC AGT CCA CCA TCT CCA GCT CGC AGA
Tyr Phe Trp Met Leu Ser Pro Phe Ser Pro Pro Ser Pro Ala Arg Arg
    1170                1175                1180

                                                                        3600
ATT ACT GAA GTT TGG GGT ATT CCT TCC CCC ATT GAT ACT GTT TTT ACT
Ile Thr Glu Val Trp Gly Ile Pro Ser Pro Ile Asp Thr Val Phe Thr
1185                1190                1195                1200

                                                    Exon IX
                                                                        3648
AGG TGC AAC TGT GAA GGA AAA ACT TTC TTC TTT AAG GAT TCT CAG TAC
Arg Cys Asn Cys Glu Gly Lys Thr Phe Phe Phe Lys Asp Ser Gln Tyr
                1205                1210                1215

                                                                        3696
TGG CGT TTT ACC AAT GAT ATA AAA GAT GCA GGG TAC CCC AAA CCA ATT
Trp Arg Phe Thr Asn Asp Ile Lys Asp Ala Gly Tyr Pro Lys Pro Ile
            1220                1225                1230
```

FIGURE 1 (Cont'd.)

```
TTC AAA GCA TTT GGA GGA CTA ACT GGA CAA ATA GTG GCA GCG CTT TCA      3744
Phe Lys Gly Phe Gly Gly Leu Thr Gly Gln Ile Val Ala Ala Leu Ser
      1235                1240                1245

ACA GCT AAA TAT AAG AAC TGG CCT GAA TCT GTG TAT TTT TTC AAG AGA      3792
Thr Ala Lys Tyr Lys Asn Trp Pro Glu Ser Val Tyr Phe Phe Lys Arg
 1250                1255                1260

          ----------Exon X----------
GGT GGC AGC ATT CAG CAG TAT ATT TAT AAA CAG GAA CCT GTA CAG AAG      3840
Gly Gly Ser Ile Gln Gln Tyr Ile Tyr Lys Gln Glu Pro Val Gln Lys
1265                1270                1275                1280

TGC CCT GGA AGA AGG CCT GCT CTA AAT TAT CCA GTG TAT GGA GAA ATG      3888
Cys Pro Gly Arg Arg Pro Ala Leu Asn Tyr Pro Val Tyr Gly Glu Met
            1285                1290                1295

ACA CAG GTT AGG AGA CGT CGC TTT GAA CGT GCT ATA GGA CCT TCT CAA      3936
Thr Gln Val Arg Arg Arg Arg Phe Glu Arg Ala Ile Gly Pro Ser Gln
      1300                1305                1310

ACA CAC ACC ATC AGA ATT CAA TAT TCA CCT GCC AGA CTG GCT TAT CAA      3984
Thr His Thr Ile Arg Ile Gln Tyr Ser Pro Ala Arg Leu Ala Tyr Gln
 1315                1320                1325

--------          Exon XI
GAC AAA GGT GTC CTT CAT AAT GAA GTT AAA GTG AGT ATA CTG TGG AGA      4032
Asp Lys Gly Val Leu His Asn Glu Val Lys Val Ser Ile Leu Trp Arg
1330                1335                1340

GGA CTT CCA AAT GTG GTT ACC TCA GCT ATA TCA CTG CCC AAC ATC AGA      4080
Gly Leu Pro Asn Val Val Thr Ser Ala Ile Ser Leu Pro Asn Ile Arg
1345                1350                1355                1360

                                                    ----------
AAA CCT GAC GGC TAT GAT TAC TAT GCC TTT TCT AAA GAT CAA TAC TAT      4128
Lys Pro Asp Gly Tyr Asp Tyr Tyr Ala Phe Ser Lys Asp Gln Tyr Tyr
            1365                1370                1375

-----Exon XII----------
AAC ATT GAT GTG CCT AGT AGA ACA GCA AGA GCA ATT ACT ACT CGT TCT      4176
Asn Ile Asp Val Pro Ser Arg Thr Ala Arg Ala Ile Thr Thr Arg Ser
      1380                1385                1390

----------
GGG CAG ACC TTA TCC AAA GTC TGG TAC AAC TGT CCT TAG ACTGATGAGC      4225
Gly Gln Thr Leu Ser Lys Val Trp Tyr Asn Cys Pro *
 1395                1400                1405

AAAGGAGGAG TCAACTAATG AAGAAATGAA TAATAAATTT TGACACTGAA              4275

AAACATTTTA TTAATAAAGA ATATTGACAT GAGTATACCA CTTTATATAT              4325

AAAAATGTTT TTAAACTTGA CAATCATTAC ACTAAAACAG ATTTGATAAT              4375

CTTATTCACA GTTGTTATTG TTTACAGACC ATTTAATTAA TATTTCCTCT              4425
```

FIGURE 1 (Cont'd.)

GTTTATTCCT CCTCTCCCTC CCATTGCATG GCTCACACCT GTAAAAGAAA 4475
AAAGAATCAA ATTGAATATA TCTTTTAAGA ATTCAAAACT AGTGTATTCA 4525
CTTACCCTAG TTCATTATAA AAAATATCTA GGCATTGTGG ATATAAAACT 4575
GTTGGGTATT CTACAACTTC AATGGAAATT ATTACAAGCA GATTAATCCC 4625
TCTTTTTGTG ACACAAGTAC AATCTAAAAG TTATATTGGA AAACATGGAA 4675
ATATTAAAAT TTTACACTTT TACTAGCTAA AACATAATCA CAAAGCTTTA 4725
TCGTGTTGTA TAAAAAAATT AACAATATAA TGGCAATAGG TAGAGATACA 4775
ACAAATGAAT ATAACACTAT AACACTTCAT ATTTTCCAAA TCTTAATTTG 4825
GATTTAAGGA AGAAATCAAT AAATATAAAA TATAAGCACA TATTTATTAT 4875
ATATCTAAGG TATACAAATC TGTCTACATG AAGTTTACAG ATTGGTAAAT 4925
ATCACCTGCT CAACATGTAA TTATTTAATA AAACTTTGGA ACATTAAAAA 4975
AATAAATTGG AGGCTTAAAA AAAAAAAAAA AAA 5008

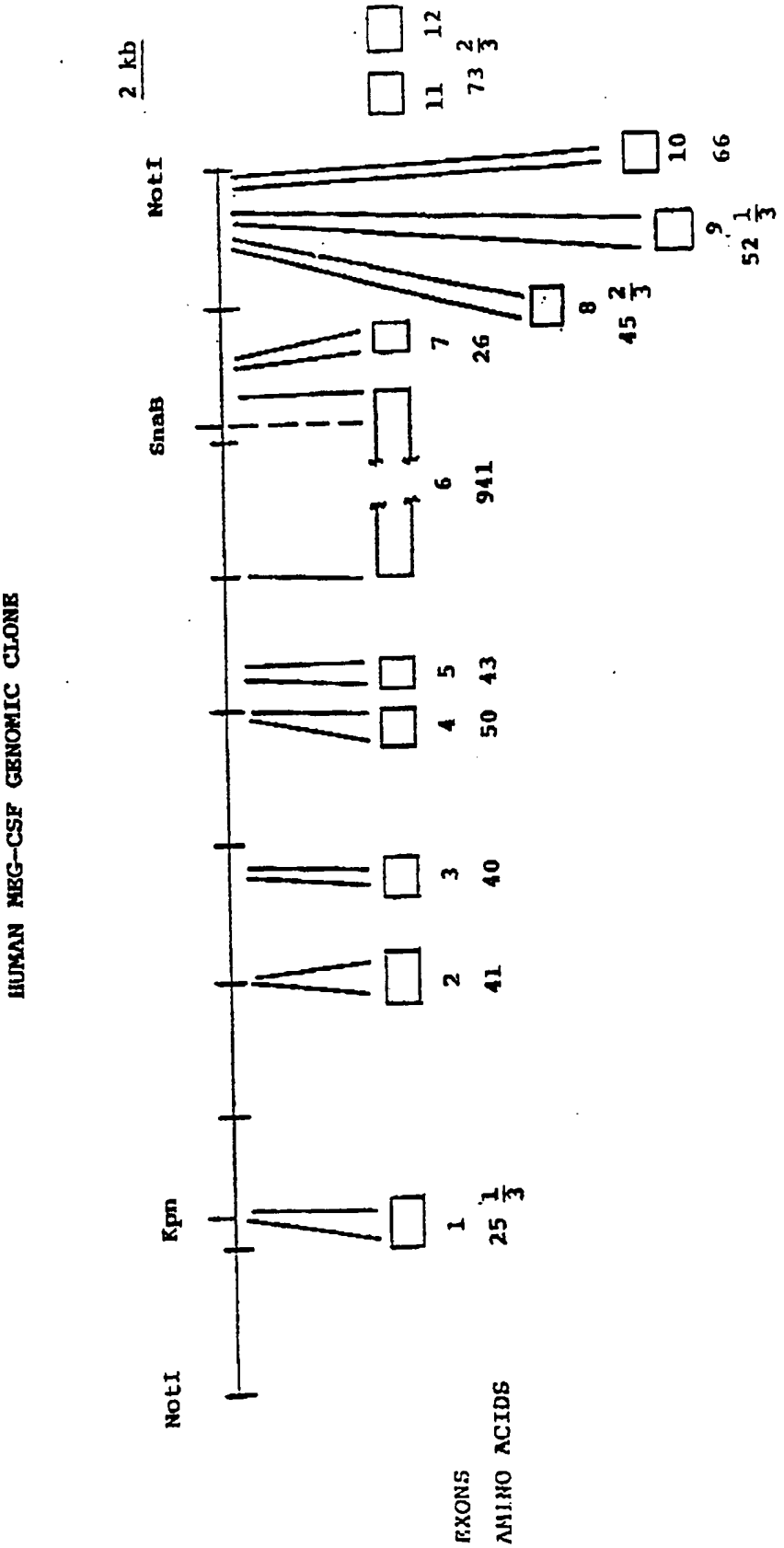
HUMAN MEG-CSF GENOMIC CLONE
2 kb
NotI
SnaB
Kpn
NotI
EXONS
AMINO ACIDS
1
25 1/3
2
41
3
40
4
50
5
43
6
941
7
26
8
45 2/3
9
52 1/3
10
66
11
73 2/3
12

FIGURE 2

FIGURE 3

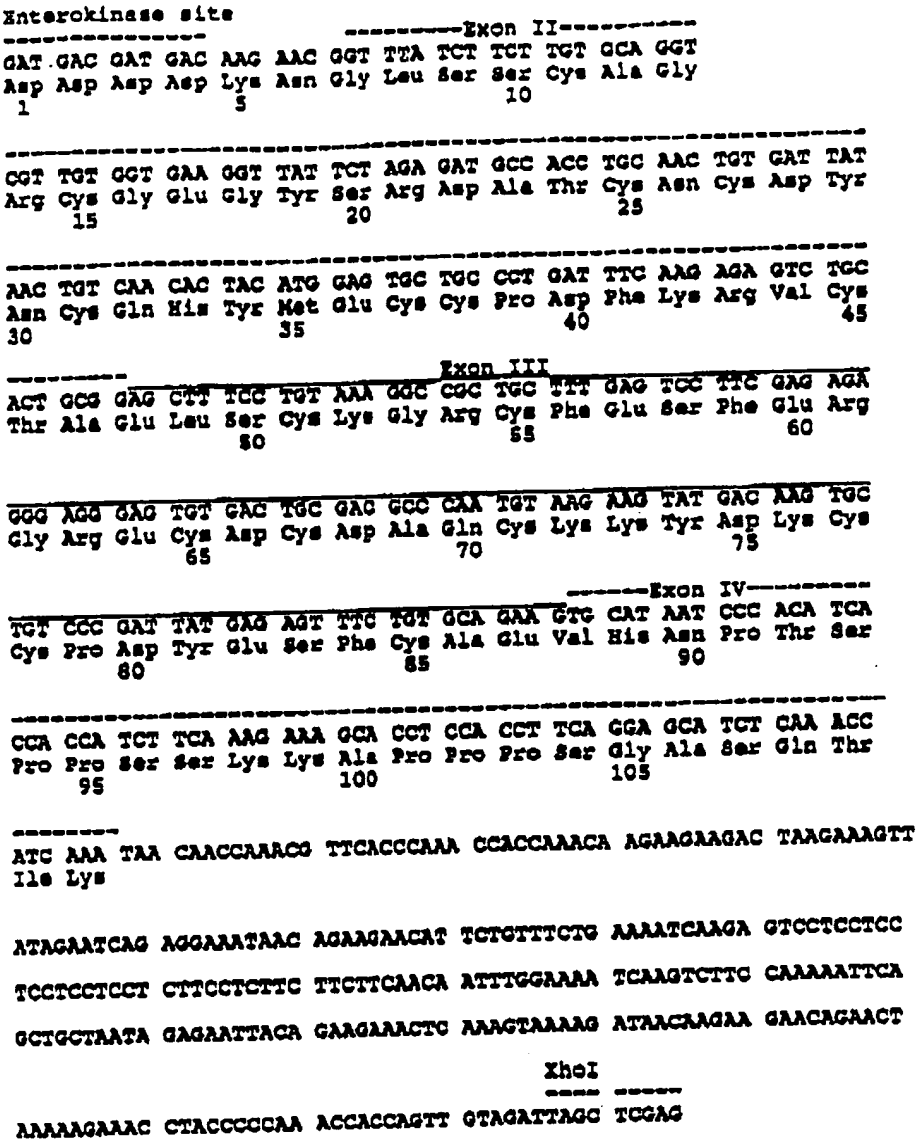
Enterokinase site
Exon II
GAT GAC GAT GAC AAG AAC GGT TTA TCT TCT TGT GCA GGT
Asp Asp Asp Asp Lys Asn Gly Leu Ser Ser Cys Ala Gly
1 5 10
CGT TGT GGT GAA GGT TAT TCT AGA GAT GCC ACC TGC AAC TGT GAT TAT
Arg Cys Gly Glu Gly Tyr Ser Arg Asp Ala Thr Cys Asn Cys Asp Tyr
15 20 25
AAC TGT CAA CAC TAC ATG GAG TGC TGC CCT GAT TTC AAG AGA GTC TGC
Asn Cys Gln His Tyr Met Glu Cys Cys Pro Asp Phe Lys Arg Val Cys
30 35 40 45
Exon III
ACT GCG GAG CTT TCC TGT AAA GGC CGC TGC TTT GAG TCC TTC GAG AGA
Thr Ala Glu Leu Ser Cys Lys Gly Arg Cys Phe Glu Ser Phe Glu Arg
50 55 60
GGG AGG GAG TGT GAC TGC GAC GCC CAA TGT AAG AAG TAT GAC AAG TGC
Gly Arg Glu Cys Asp Cys Asp Ala Gln Cys Lys Lys Tyr Asp Lys Cys
65 70 75
Exon IV
TGT CCC GAT TAT GAG AGT TTC TGT GCA GAA GTG CAT AAT CCC ACA TCA
Cys Pro Asp Tyr Glu Ser Phe Cys Ala Glu Val His Asn Pro Thr Ser
80 85 90
CCA CCA TCT TCA AAG AAA GCA CCT CCA CCT TCA GGA GCA TCT CAA ACC
Pro Pro Ser Ser Lys Lys Ala Pro Pro Pro Ser Gly Ala Ser Gln Thr
95 100 105
ATC AAA TAA CAACCAAACG TTCACCCAAA CCACCAAACA AGAAGAAGAC TAAGAAAGTT
Ile Lys
ATAGAATCAG AGGAAATAAC AGAAGAACAT TCTGTTTCTG AAAATCAAGA GTCCTCCTCC
TCCTCCTCCT CTTCCTCTTC TTCTTCAACA ATTTGGAAAA TCAAGTCTTC CAAAAATTCA
GCTGCTAATA GAGAATTACA GAAGAAACTC AAAGTAAAAG ATAACAAGAA GAACAGAACT
XhoI
AAAAAGAAAC CTACCCCCAA ACCACCAGTT GTAGATTAGC TCGAG

FIGURE 4

```
GACGAAAGGG CCTCGTGATA CGCCTATTTT TATAGGTTAA          40
TGTCATGATA ATAATGGTTT CTTAGACGTC AGGTGGCACT          80
TTTCGGGGAA ATGTGCGCGG AACCCCTATT TGTTTATTTT         120
TCTAAATACA TTCAAATATG TATCCGCTCA TGAGACAATA         160
ACCCTGATAA ATGCTTCAAT AATATTGAAA AAGGAAGAGT         200
ATGAGTATTC AACATTTCCG TGTCGCCCTT ATTCCCTTTT         240
TTGCGGCATT TTGCCTTCCT GTTTTTGCTC ACCCAGAAAC         280
GCTGGTGAAA GTAAAAGATG CTGAAGATCA GTTGGGTGCA         320
CGAGTGGGTT ACATCGAACT GGATCTCAAC AGCGGTAAGA         360
TCCTTGAGAG TTTTCGCCCC GAAGAACGTT TTCCAATGAT         400
GAGCACTTTT AAAGTTCTGC TATGTGGCGC GGTATTATCC         440
CGTATTGACG CCGGGCAAGA GCAACTCGGT CGCCGCATAC         480
ACTATTCTCA GAATGACTTG GTTGAGTACT CACCAGTCAC         520
AGAAAAGCAT CTTACGGATG GCATGACAGT AAGAGAATTA         560
TGCAGTGCTG CCATAACCAT GAGTGATAAC ACTGCGGCCA         600
ACTTACTTCT GACAACGATC GGAGGACCGA AGGAGCTAAC         640
CGCTTTTTTG CACAACATGG GGGATCATGT AACTCGCCTT         680
GATCGTTGGG AACCGGAGCT GAATGAAGCC ATACCAAACG         720
ACGAGCGTGA CACCACGATG CCTGTAGCAA TGGCAACAAC         760
GTTGCGCAAA CTATTAACTG GCGAACTACT TACTCTAGCT         800
TCCCGGCAAC AATTAATAGA CTGGATGGAG GCGGATAAAG         840
TTGCAGGACC ACTTCTGCGC TCGGCCCTTC CGGCTGGCTG         880
GTTTATTGCT GATAAATCTG GAGCCGGTGA GCGTGGGTCT         920
CGCGGTATCA TTGCAGCACT GGGGCCAGAT GGTAAGCCCT         960
CCCGTATCGT AGTTATCTAC ACGACGGGGA GTCAGGCAAC        1000
```

FIGURE 4 (CONT'D.)

TATGGATGAA CGAAATAGAC AGATCGCTGA GATAGGTGCC 1040

TCACTGATTA AGCATTGGTA ACTGTCAGAC CAAGTTTACT 1080

CATATATACT TTAGATTGAT TTAAAACTTC ATTTTTAATT 1120

TAAAAGGATC TAGGTGAAGA TCCTTTTTGA TAATCTCATG 1160

ACCAAAATCC CTTAACGTGA GTTTTCGTTC CACTGAGCGT 1200

CAGACCCCGT AGAAAAGATC AAAGGATCTT CTTGAGATCC 1240

TTTTTTTCTG CGCGTAATCT GCTGCTTGCA AACAAAAAAA 1280

CCACCGCTAC CAGCGGTGGT TTGTTTGCCG GATCAAGAGC 1320

TACCAACTCT TTTTCCGAAG GTAACTGGCT TCAGCAGAGC 1360

GCAGATACCA AATACTGTCC TTCTAGTGTA GCCGTAGTTA 1400

GGCCACCACT TCAAGAACTC TGTAGCACCG CCTACATACC 1440

TCGCTCTGCT AATCCTGTTA CCAGTGGCTG CTGCCAGTGG 1480

CGATAAGTCG TGTCTTACCG GGTTGGACTC AAGACGATAG 1520

TTACCGGATA AGGCGCAGCG GTCGGGCTGA ACGGGGGGTT 1560

CGTGCACACA GCCCAGCTTG GAGCGAACGA CCTACACCGA 1600

ACTGAGATAC CTACAGCGTG AGCATTGAGA AAGCGCCACG 1640

CTTCCCGAAG GGAGAAAGGC GGACAGGTAT CCGGTAAGCG 1680

GCAGGGTCGG AACAGGAGAG CGCACGAGGG AGCTTCCAGG 1720

GGGAAACGCC TGGTATCTTT ATAGTCCTGT CGGGTTTCGC 1760

CACCTCTGAC TTGAGCGTCG ATTTTTGTGA TGCTCGTCAG 1800

GGGGGCGGAG CCTATGGAAA AACGCCAGCA ACGCGGCCTT 1840

TTTACGGTTC CTGGCCTTTT GCTGGCCTTT TGCTCACATG 1880

TTCTTTCCTG CGTTATCCCC TGATTCTGTG GATAACCGTA 1920

TTACCGCCTT TGAGTGAGCT GATACCGCTC GCCGCAGCCG 1960

AACGACCGAG CGCAGCGAGT CAGTGAGCGA GGAAGCGGAA 2000

GAGCGCCCAA TACGCAAACC GCCTCTCCCC GCGCGTTGGC 2040

FIGURE 4 (CONT'D)

```
CGATTCATTA ATGCAGAATT GATCTCTCAC CTACCAAACA                    2080

ATGCCCCCCT GCAAAAAATA AATTCATATA AAAAACATAC                    2120

AGATAACCAT CTGCGGTGAT AAATTATCTC TGGCGGTGTT                    2160

GACATAAATA CCACTGGCGG TGATACTGAG CACATCAGCA                    2200

GGACGCACTG ACCACCATGA ATTCAAGAAG GAGATATACA                    2240

T ATG AGC GAT AAA ATT ATT CAC CTG ACT GAC GAC                  2274
  Met Ser Asp Lys Ile Ile His Leu Thr Asp Asp
    1               5                  10

AGT TTT GAC ACG GAT GTA CTC AAA GCG GAC GGG                    2307
Ser Phe Asp Thr Asp Val Leu Lys Ala Asp Gly
               15                  20

GCG ATC CTC GTC GAT TTC TGG GCA GAG TGG TGC                    2340
Ala Ile Leu Val Asp Phe Trp Ala Glu Trp Cys
          25                  30

GGT CCG TGC AAA ATG ATC GCC CCG ATT CTG GAT                    2373
Gly Pro Cys Lys Met Ile Ala Pro Ile Leu Asp
     35                  40

GAA ATC GCT GAC GAA TAT CAG GGC AAA CTG ACC                    2406
Glu Ile Ala Asp Glu Tyr Gln Gly Lys Leu Thr
 45                  50                  55

GTT GCA AAA CTG AAC ATC GAT CAA AAC CCT GGC                    2439
Val Ala Lys Leu Asn Ile Asp Gln Asn Pro Gly
                60                  65

ACT GCG CCG AAA TAT GGC ATC CGT GGT ATC CCG                    2472
Thr Ala Pro Lys Tyr Gly Ile Arg Gly Ile Pro
               70                  75

ACT CTG CTG CTG TTC AAA AAC GGT GAA GTG GCG                    2505
Thr Leu Leu Leu Phe Lys Asn Gly Glu Val Ala
          80                  85

GCA ACC AAA GTG GGT GCA CTG TCT AAA GGT CAG                    2538
Ala Thr Lys Val Gly Ala Leu Ser Lys Gly Gln
     90                  95

TTG AAA GAG TTC CTC GAC GCT AAC CTG GCC GGT                    2571
Leu Lys Glu Phe Leu Asp Ala Asn Leu Ala Gly
100                 105                 110
```

FIGURE 4 (CONT'D)

```
TCT GGT TCT GGT GAT GAC GAT GAC AAA GGT CCA                    2604
Ser Gly Ser Gly Asp Asp Asp Asp Lys Gly Pro
                115                     120

CCA CCA GGT CCA CCT CGA GTT TCC CCA GAC CCT                    2637
Pro Pro Gly Pro Pro Arg Val Ser Pro Asp Pro
            125                 130

CGG GCC GAG CTG GAC AGC ACC GTG CTC CTG ACC                    2670
Arg Ala Glu Leu Asp Ser Thr Val Leu Leu Thr
        135                 140

CGC TCT CTC CTG GCG GAC ACG CGG CAG CTG GCT                    2703
Arg Ser Leu Leu Ala Asp Thr Arg Gln Leu Ala
    145                 150

GCA CAG CTG AGG GAC AAA TTC CCA GCT GAC GGG                    2736
Ala Gln Leu Arg Asp Lys Phe Pro Ala Asp Gly
155                 160                 165

GAC CAC AAC CTG GAT TCC CTG CCC ACC CTG GCC                    2769
Asp His Asn Leu Asp Ser Leu Pro Thr Leu Ala
                170                 175

ATG AGT GCG GGG GCA CTG GGA GCT CTA CAG CTC                    2802
Met Ser Ala Gly Ala Leu Gly Ala Leu Gln Leu
            180                 185

CCA GGT GTG CTG ACA AGG CTG CGA GCG GAC CTA                    2835
Pro Gly Val Leu Thr Arg Leu Arg Ala Asp Leu
        190                 195

CTG TCC TAC CTG CGG CAC GTG CAG TGG CTG CGC                    2868
Leu Ser Tyr Leu Arg His Val Gln Trp Leu Arg
    200                 205

CGG GCA GGT GGC TCT TCC CTG AAG ACC CTG GAG                    2901
Arg Ala Gly Gly Ser Ser Leu Lys Thr Leu Glu
210                 215                 220

CCC GAG CTG GGC ACC CTG CAG GCC CGA CTG GAC                    2934
Pro Glu Leu Gly Thr Leu Gln Ala Arg Leu Asp
                225                 230

CGG CTG CTG CGC CGG CTG CAG CTC CTG ATG TCC                    2967
Arg Leu Leu Arg Arg Leu Gln Leu Leu Met Ser
            235                 240
```

FIGURE 4 (CONT'D)

```
CGC CTG GCC CTG CCC CAG CCA CCC CCG GAC CCG                    3000
Arg Leu Ala Leu Pro Gln Pro Pro Pro Asp Pro
        245                 250

CCG GCG CCC CCG CTG GCG CCC CCC TCC TCA GCC                    3033
Pro Ala Pro Pro Leu Ala Pro Pro Ser Ser Ala
    255                 260

TGG GGG GGC ATC AGG GCC GCC CAC GCC ATC CTG                    3066
Trp Gly Gly Ile Arg Ala Ala His Ala Ile Leu
265                 270                 275

GGG GGG CTG CAC CTG ACA CTT GAC TGG GCC GTG                    3099
Gly GLy Leu His Leu Thr Leu Asp Trp Ala Val
                280                 285

AGG GGA CTG CTG CTG CTG AAG ACT CGG CTG TGA                    3132
Arg Gly Leu Leu Leu Leu Lys Thr Arg Leu
            290                 295

AAGCTTATCG ATACCGTCGA CCTGCAGTAA TCGTACAGGG                    3172

TAGTACAAAT AAAAAAGGCA CGTCAGATGA CGTGCCTTTT                    3212

TTCTTGTGAG CAGTAAGCTT GGCACTGGCC GTCGTTTTAC                    3252

AACGTCGTGA CTGGGAAAAC CCTGGCGTTA CCCAACTTAA                    3292

TCGCCTTGCA GCACATCCCC CTTTCGCCAG CTGGCGTAAT                    3332

AGCGAAGAGG CCCGCACCGA TCGCCCTTCC CAACAGTTGC                    3372

GCAGCCTGAA TGGCGAATGG CGCCTGATGC GGTATTTTCT                    3412

CCTTACGCAT CTGTGCGGTA TTTCACACCG CATATATGGT                    3452

GCACTCTCAG TACAATCTGC TCTGATGCCG CATAGTTAAG                    3492

CCAGCCCCGA CACCCGCCAA CACCCGCTGA CGCGCCCTGA                    3532

CGGGCTTGTC TGCTCCCGGC ATCCGCTTAC AGACAAGCTG                    3572

TGACCGTCTC CGGGAGCTGC ATGTGTCAGA GGTTTTCACC                    3612

GTCATCACCG AAACGCGCGA                                         3632
```

INTERNATIONAL SEARCH REPORT

International Application No PCT/US 92/00433

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all)*

According to International Patent Classification (IPC) or to both National Classification and IPC

Int.Cl. 5 C12N15/27; C07K13/00; A61K37/02

II. FIELDS SEARCHED

Minimum Documentation Searched[7]

Classification System	Classification Symbols
Int.Cl. 5	C07K

Documentation Searched other than Minimum Documentation to the Extent that such Documents are included in the Fields Searched[8]

III. DOCUMENTS CONSIDERED TO BE RELEVANT[9]

Category *	Citation of Document, [11] with indication, where appropriate, of the relevant passages [12]	Relevant to Claim No.[13]
X	THE JOURNAL OF CLINICAL INVESTIGATION vol. 75, no. 4, April 1985, THE AMERICAN SOCIETY FOR CLIN. INV. pages 1174 - 1182; HOFFMAN, R. ET AL.: 'Purification and partial characterization of a megakaryocyte colony-stimulating factor from human plasma.' cited in the application See page 1180, "Discussion"; paragraphs 3,4,5; page 1181, paragraph 2 ---	1-10
X	EXPERIMENTAL HEMATOLOGY vol. 14, no. 6, 1986, NEW YORK, USA page 490; SHIMIZU, T. ET AL.: 'Purification of human megakaryocyte colony stimulating factor.' see the whole document --- -/--	1-10

* Special categories of cited documents : [10]

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

IV. CERTIFICATION

Date of the Actual Completion of the International Search	Date of Mailing of this International Search Report
23 JUNE 1992	10. 07. 92
International Searching Authority	**Signature of Authorized Officer**
EUROPEAN PATENT OFFICE	NAUCHE S.A.

Form PCT/ISA/210 (second sheet) (January 1985)

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, with indication, where appropriate, of the relevant passages	Relevant to Claim No.
X	EXPERIMENTAL HEMATOLOGY vol. 15, no. 5, 1987, NEW YORK, USA page 492; SHIMIZU, T.: 'Purification of human sialylated megakaryocyte colony stimulating factor.' see the whole document	1-10
A	EXPERIMENTAL HEMATOLOGY vol. 13, no. 11, December 1985, NEW YORK, USA pages 1164 - 1172; MAZUR, E.M. ET AL.: 'Human megakaryocyte colony stimulating factor in sera from aplastic dogs : partial purification, characterization and determination of hematopoietic cell lineage specificty.' cited in the application	
P,X	WO,A,9 012 108 (GENETICS INSTITUTE) 18 October 1990 see claims 1-20	1-10
P,X	WO,A,9 102 001 (GENETICS INSTITUTE) 21 February 1991 see the claims	1-10

ANNEX TO THE INTERNATIONAL SEARCH REPORT
ON INTERNATIONAL PATENT APPLICATION NO. US 9200433
SA 57766

This annex lists the patent family members relating to the patent documents cited in the above-mentioned international search report.
The members are as contained in the European Patent Office EDP file on
The European Patent Office is in no way liable for these particulars which are merely given for the purpose of information. 23/06/92

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9012108	18-10-90	AU-A- 5405890	05-11-90
		EP-A- 0466780	22-01-92
WO-A-9102001	21-02-91	AU-A- 6295790	11-03-91
		EP-A- 0487613	03-06-92

EPO FORM P0479

For more details about this annex : see Official Journal of the European Patent Office, No. 12/82